Crystal structure of infectious bursal disease virus VP2 subviral particle at 2.6 Å resolution: Implications in virion assembly and immunogenicity

Cheng-Chung Lee a,b,1, Tzu-Ping Ko b,1, Chia-Cheng Chou c, Masato Yoshimura d, Shyue-Ru Doong e, Min-Ying Wang f,*, Andrew H.-J. Wang b,c,*

a Structural Biology Program, Institute of Biochemistry and Molecular Biology, National Yang-Ming University, Taipei 11221, Taiwan, ROC
b Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan, ROC
c National Core Facility of High-Throughput Protein Crystallography, Academia Sinica, Taipei 11529, Taiwan, ROC
d Institute for Protein Research, Osaka University, Osaka 565-0871, Japan
e Department of Chemical and Materials Engineering, National Central University, Taoyuan 32001, Taiwan, ROC
f Graduate Institute of Biotechnology, National Chung-Hsing University, Taichung 40227, Taiwan, ROC

Received 3 January 2006; received in revised form 27 February 2006; accepted 28 February 2006
Available online 31 March 2006

Abstract

The structural protein VP2 of infectious bursal disease virus (IBDV) spontaneously forms a dodecahedral T = 1 subviral particle (SVP), and is a primary immunogen of the virus. In this study, the structure of IBDV SVP was determined in a cubic crystal and refined to 2.6 Å resolution. It contains 20 independent VP2 subunits in a crystallographic asymmetric unit. Each subunit is folded mainly into a shell domain and a protrusion domain, both with the Swiss-roll topology, plus a small helical base domain. Three VP2 subunits constitute a tight trimer, which is the building block of IBDV (sub)viral particles. The structure revealed a calcium ion bound to three pairs of symmetry-related Asp31 and Asp174 to stabilize the VP2 trimer. Our results of treatment of SVP with EGTA, a Ca2+-chelating reagent, indicated that the metal-ion may be important not only in maintaining highly stable quaternary structure but also in regulating the swelling and dissociation of the icosahedral particles. A Ca2+-dependent assembly pathway was thus proposed, which involves further interactions between the trimers. The 20 independent subunits showed conformational variations, with the surface loops of the protrusion domain being the most diverse. These loops are targets of the neutralizing antibodies. Several common interactions between the surface loops were clearly observed, suggesting a possible major conformation of the immunogenic epitopes.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Calcium ion; Molecular interactions; Icosahedral symmetry; Viral swelling; Epitope

1. Introduction

Infectious bursal disease virus (IBDV)2 has been causing immunosuppressive viral disease in avians and leading to serious loss in the poultry industry (Muller et al., 2003). IBDV belongs to Avibirnavirus genus of the Birnaviridae family. The viral genome comprises two segments, A and B, of double-stranded RNA (Dobos et al., 1979). The viral
proteins VP2 and VP3 encoded by segment A constitute the unenveloped T = 13 icosahedral capsid, which has a diameter of about 700 Å (Bottcher et al., 1997). VP2 is the primary immunogen (Heine et al., 1991). When VP2 was expressed in baculovirus alone, it spontaneously formed a dodecahedral T = 1 subviral particle (SVP) that showed similar immunogenicity as the intact IBDV virion (Caston et al., 2001). Based on cryo-electron microscopic structures, the antigenic determinants have been supposed to associate with the protrusions at the triad of the virus particles. Two conformational epitopes in the regions of 210–225 and 312–324 of VP2 sequence were found important in the recognition by antibodies (Heine et al., 1991). The crystal structures of both T = 13 and T = 1 particles revealed that the single shell virial capsids comprise of VP2 trimers (Coulibaly et al., 2005). Each polypeptide folds into a helical base (B) domain, a shell (S) domain, and a protrusion (P) domain, with the latter two resembling the β-barrel domains of nodavirus and rotavirus, respectively. The two epitope regions are located in the outermost surface loop of the protrusion domain.

In addition to epitope recognition by antibodies, the hypervariable region, i.e., amino acid 204–344 of VP2, is presumably also responsible for the interaction with the cellular receptors and restriction in infectivity (Van Loon et al., 2002). Two different serotypes of IBDV (serotype 1 and 2) have been described. Serotype 1 IBDV strains are pathogenic only in chicken and specifically infect developing B-lymphoid cells in the bursa of Fabricius, whereas serotype 2 are found also in other fowls and able to replicate naturally in different tissues of bird (Muller et al., 2003). Serotype 1 isolates are subdivided into classical, antigenic variant, and very virulent strains. Amino acid substitutions in the hypervariable region of VP2 different strains appeared to be associated with altered antigenicity and virulence. To precisely explain these biological functions and interactions, a high-resolution structure is required.

As reported by Coulibaly et al. (2005), SVPs are exclusively formed by VP2 and contain no nucleic acids or other heterologous proteins, and they are significantly different from the IBDV virion. Thus, the assembly of SVPs mainly involves the interaction of homogenous VP2. Although it appears to be a simple pathway, no information regarding SVP assembly has been available. Our unpublished data showed that the assembly efficiency of SVPs formed by rVP2H (VP2 protein with 452 amino acid and 6 extra histidine residues) is about 90%. This extremely high efficiency of particulate assembly indicates that other factors (for example, metal ions) may be involved in the subunit interface. Previously, we crystallized IBDV SVP formed by rVP2H in a cubic unit cell (Lee et al., 2003). The diffraction data at 6 Å resolution showed strong icosahedral symmetry. We predicted that the cubic crystal might contain four SVPs in a unit cell, and each asymmetric unit would comprise one-third of the particle. Improvement of the crystallization conditions yielded better crystals that diffracted X-rays to well beyond 3 Å resolution. In this paper, we report the cubic crystal structure solved by molecular replacement using the PDB coordinates of 1WCD (Coulibaly et al., 2005). Subsequent refinement at 2.6 Å resolution yielded a model of 20 protein subunits, in which structural heterogeneity was observed, as well as some other new features including cis-peptide bonds and bound ions. By analyzing the subunit interactions, we also propose a mechanism for the self-assembly of IBDV VP2 into SVP. Our high-resolution structure allowed us to locate a calcium ion on the triad axis, which was then shown to play an important role in the integrity of the SVP.

It has been demonstrated that removal of the zinc ion destabilizes the VP6 trimer of rotavirus (Erk et al., 2003). Two recent studies showed that the assembly of human polyomavirus BK is dependent on calcium (Nilsson et al., 2005), and the swelling of tomato busy stunt virus is dependent on divalent ions (Aramayo et al., 2005). Whether bound at the triad or not, metal ions appear to play a vital role in the assembly of many viral capsids. Besides, Ca$^{2+}$ is important in the life cycle and for the stability of several isosahedral viruses. In rotavirus, virus entry, activation of transcription, morphogenesis, cell lysis, particle release, and the distant action of viral proteins are Ca$^{2+}$ dependent processes (Ruiz et al., 2000). For the simian virus SV40, calcium ion mediates not only virion assembly but also the initial infection processes of cell entry and nuclear entry (Li et al., 2003). To investigate the effect of Ca$^{2+}$ on the IBDV SVP structure, we used EGTA to remove the divalent ion and analyzed the particle morphology by gel electrophoresis and electron microscopy.

2. Materials and methods

2.1. Crystallization and data collection

Crystals of IBDV SVP were obtained as described previously (Lee et al., 2003). The recombinant VP2 protein that contained a C-terminal His$_{6}$-tag (rVP2H, 458 amino acids) was expressed in High-Five cells, purified by immobilized metal-ion chromatography with a Ni–NTA column, and concentrated to 8 mg/ml. For sitting drop crystallization, 2 μl of the rVP2H solution was mixed with 2 μl of a reservoir that contained 0.1 M MES (2-(N-morpholino)-ethanesulfonic acid) pH 6.5 and 12% PEG20000 (polyethylene glycol) as a precipitant, and equilibrated with the reservoir. All diffraction data were collected at 100 K. Prior to flash-cooling, crystals were rinsed with a cryoprotectant solution of 65% MPD (2-methyl-2,4-pentanediol) and 35% reservoir. X-ray diffraction data from the native crystal were collected to 2.6 Å resolution at wavelength λ = 1.000 Å using the synchrotron beam line AR-NW12 of Photon Factory, Tsukuba, Japan. The platinum derivative crystals were obtained by mixing 2 μl rVP2H solution with 2 μl of the same reservoir with an additional 2.5 μM di-m-iodo-bis(ethylenediamine) diplatinum (PIP). Data were collected at λ = 1.071 Å, using beam line BL17B2 at National Synchrotron Radiation Research Center, Hsinchu, Taiwan. The derivative crystals...
of Ta₆Br₁₂⁺ cluster were prepared by soaking native crystals in a cryoprotectant solution that contained the heavy-atom compound of Ta₆Br₁₄ at a concentration of 2.5 mM for 4 days, after which the colorless pyramid-shaped crystals became green. Data were collected at λ = 1.2530 Å, corresponding to the L III absorption energy of tantalum, using beam line BL18B at the Photon Factory. All diffraction images were recorded using CCD detectors (charged-coupled device) and data processed using the software package of DENZO and SCALEPACK (Otwinowski and Minor, 1997). Statistics are shown in Table 1.

2.2. Computer programs

Structure determination and computational refinement employed the software of Crystallography and NMR System (CNS; Brünger et al., 1998), in which 5% of randomly selected data were set aside to calculate R_free value as a monitor (Brünger, 1993). Manual building of the model was carried out using the program O (Jones et al., 1991). For structural analysis, the software package of CCP4 suite (Collaborative Computational Project Number 4, 1994) and the program GRASP (Nicholls et al., 1991) were also used. Figures were prepared using the programs of PyMOL (DeLano, 2002), MolScript (Kraulis, 1991) and Raster3D (Merritt and Murphy, 1994).

2.3. Structure determination and refinement

2.3.1. Molecular replacement

The cubic crystal structure of IBDV SVP was solved by molecular replacement using the model of PDB 1WCD as a search model. The space group has been determined as P2₁3 (Lee et al., 2003). A 60-subunit particle was generated, in which the threefold axis was parallel to the z-axis (as derived from the hexagonal P6₃ crystal of 1WCD) and the center of mass was replaced at the origin. The rotation function was calculated using CNS (Brünger et al., 1998), with more than 25 000 reflections of the native data set in the resolution range of 18–6 Å. The search was carried out with two angles fixed (θ₂ = 54.74° and θ₃ = 45°), and allowing only one-dimensional rotation (θ₁) about the triad, along the body diagonal of the cubic unit cell. The function showed a sharp peak with about 9.5° rotation (Supplementary Figure S1). The translation function also yielded a single peak, with the center of SVP remained at the origin of the unit cell.

The model was then reduced to an asymmetric unit, which contained 20 subunits of the VP2 protein, comprising 1/3 of the SVP. Rigid body refinement to 2.6 Å yielded an R value of 0.32, and it decreased to 0.26 after preliminary energy minimization and B value refinements. The ini-

Table 1

| Data collection and refinement statistics of the IBDV crystals a |
|-----------------------------|-----------------|-----------------|
| Data collection              | Native          | PIP b           | Ta₆Br₁₂⁺ b       |
| Space group                  | P2₁3            | P2₁3            | P2₁3             |
| Wavelength (Å)               | 1.000           | 1.072           | 1.253            |
| Unit cell dimension (Å)      | 316.4           | 316.0           | 314.8            |
| Resolution range (Å)         | 40–2.6 (2.69–2.60) | 99–2.7 (2.79–2.70) | 50–3.5 (3.63–3.50) |
| Number of observations       | 1 449 380 (90 396) | 191 816 (191 178) | 513 417 (193 868) |
| Unique reflections           | 318 732 (30 571) | 283 947 (26 800) | 122 861 (55 958) |
| Completeness (%)             | 99.5 (96.0)     | 99.3 (94.4)     | 94.4 (74.6)      |
| Average I/σ(I)               | 18.6 (2.1)      | 12.3 (1.5)      | 7.9 (1.8)        |
| Rmerge (%)                   | 9.6 (57.8)      | 12.3 (68.6)     | 14.0 (50.8)      |

<table>
<thead>
<tr>
<th>Refinement</th>
<th>Merge (native + PIP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>40–2.6 (2.69–2.60)</td>
</tr>
<tr>
<td>Number of positive reflections (F &gt; 0)</td>
<td>286 954 (23 504)</td>
</tr>
<tr>
<td>R_cryst (%) for 95% data</td>
<td>16.5 (24.9)</td>
</tr>
<tr>
<td>R_free (%) for 5% data</td>
<td>21.5 (30.3)</td>
</tr>
<tr>
<td>Root-mean-square deviation from ideal bond lengths (Å)</td>
<td>0.016</td>
</tr>
<tr>
<td>Root-mean-square deviation from ideal bond angles (°)</td>
<td>1.9</td>
</tr>
</tbody>
</table>

Ramachandran plot: number of non-proline, non-glycine, and non-terminal residues

| 6221 in most favored regions (%) | 88.7 |
| 772 in additional allowed regions (%) | 11.0 |
| 0 in generously allowed regions (%) | 0.0 |
| 20 in disallowed regions (%) | 0.3 |

Average B value (Å²)

| 33 160 back-bone atoms | 40.6 |
| 29 335 side-chain atoms | 42.2 |
| 4 445 water molecules | 49.7 |
| 16 ions (4 at the triad) | 34.5 |

a Numbers in parentheses are for the outmost resolution shell.
b Data sets were collected at the peak wavelengths and processed for anomalous signals.
tial $2F_O - F_C$ map showed clear tracing of most polypeptide chains, except the region of 113–120 near the fivefold axis that lacked strong densities. Non-crystallographic symmetry (NCS) averaging and density modification did not improve the map. The 9.5° rotation resulted in non-identical environment of each monomer, and thus led to slight variation of the subunit conformation. In particular, some loops in the projection domains clashed with surface residues in neighboring particles. However, there were clear electron densities corresponding to the altered loop structures in these regions, which allowed relocation of the polypeptide chains.

2.3.2. Model building

An initial model was built according to the sequence of P3009 strain (GenBank Accession No. AF109154), by visual inspection of the electron density map with the program O. In our complete model, three residues are different from the sequence of P3009. For consistency, the residue 135 should be a Glu not Asp, and residue 330 and 331 are Ser and Gly, but not Met and Trp, because the electron density maps showed clear features of those residues. The changes were later confirmed by DNA sequencing.

The refinement continued with manual modifications of the model and computational adjustments of the coordinates and temperature factors. Equivalent regions in the 20 subunits that did not show significant deviations were treated with strong NCS restraints. Addition of 10 mM CaCl$_2$ and incubation at room temperature for 4 h. These protein solutions (about 1 mg/ml) were analyzed using 1% agarose gel electrophoresis.

2.3.3. Heavy atom sites

The partially refined models were exploited to locate heavy atom sites in the derivative crystals by calculating anomalous difference Fourier maps using the structure factors of $\Delta F_{\text{HA}}$ and phase angles of native $\Phi_{\text{NAT}} + 90^\circ$. The data set of PIP derivative yielded a featureless map, which did not show any prominent peak for platinum site. Because the unit cell dimension of the PIP derivative was very similar to the native crystal, 260369 Friederich pairs of the PIP data set were averaged with an $R_{\text{anomalous}}$ of 0.059. The new data set was then merged with the native data set, giving an overall $R_{\text{merge}}$ of 0.113 for 297238 matched reflections, and these data with significantly higher redundancy were used in the subsequent refinement. On the other hand, the derivative of Ta$_6$Br$_{12}^{2+}$ showed a major site near the triad axis, at the interface of neighboring particles (Supplemental Figure S2). The site had low occupancy, thus insufficient for phase determination. This data set was not used in refinement. The refinement procedure is summarized in Supplemental Table S1 and the final model statistics in Table 1.

2.4. Sequence comparison

Many sequences for different strains of IBDV are available from the National Center for Biotechnology Information (NCBI) and categorized in two serotypes. We chose to compare the hypervariable region (residues 173–341) of rVP2H sequence with those of the USA classic strain STC, the very virulent strain HK46 of serotype 1 and the OH strain of serotype 2, which also infects duck and turkey. The amino acid sequences of STC, HK46, and OH strains were obtained from NCBI with Accession Nos. P22351, AAD23373, and P27276, respectively.

2.5. Determination of metal content by ICP-MS

The metal content of SVP was estimated by an inductively coupled plasma mass spectrometer (ICP-MS, SCIEX ELAN 5000, Perkin Elmer, USA). The purified SVP was dialyzed in 50 mM Tris buffer (pH 7.0) and 1 ml solution (1 mg/ml) was used.

2.6. Gel electrophoresis

IBDV SVP solutions were first dialyzed against 50 mM Tris buffers of two different pH values (7.0 and 8.5). After addition of 100 mM EGTA at room temperature, the samples were incubated overnight, and dialyzed again 50 mM Tris buffer (pH 7.0 and 8.5) to remove EGTA and calcium ion. The EGTA and calcium-free samples were then treated with 0.10 M CaCl$_2$ and incubated at room temperature for 4 h. These protein solutions (about 8 μg each) were analyzed using 1% agarose gel electrophoresis, carried out at 50 V and 4°C for 2–3 h in 50 mM Tris-acetate buffer (pH 7.0 and 8.5). The gels were stained with Instant Blue (GeneMarK Technology Co., Ltd, Taiwan).

2.7. Electron microscopy

Standard negative staining procedures were used to prepare the VP2 particles for electron microscopic analysis. The samples of 5 μl were placed on Formvar–carbon-coated copper grids and allowed to settle for 30 s. The solution was removed by blotting with a filter paper and the grid stained with 2% aqueous uranium acetate solution for 1 min, then the uranium acetate solution was also removed. Electron micrographs were recorded using a Jeol (Japan Electron Optics Laboratory) JEM 1200 EX-2 transmission electron microscope operating at 120 kV.
3. Results

3.1. Monomer VP2 structure

The refined model of IBDV SVP contains 20 independent polypeptide chains, which are designated subunits A–T. The N-termini start at residue 11 or 12, and the C-termini stop at residues 427–430, depending on the electron densities for individual subunits. The first 10 and the last 28 residues including the His-tag were not observed, probably because they were highly flexible. Another disordered region was located in residues 113–120. Only four subunits (A, B, L, and T) had densities for building continuous polypeptide chains. Each subunit folds into two major “Swiss roll” β-barrels and three α-helices, as shown in Fig. 1. The α-helices are associated to form the base (B) domain. The first β-barrel encompasses amino acid residues 35–173 and 348–376; the second barrel residues 202–341. They constitute the shell (S) domain and the protrusion (P) domain, respectively, and are disposed roughly perpendicular to each other. Beyond strand C of the S domain (SC) there are two peripheral strands SC and SCw (residues 65–80). On the other side there is another small strand of Sf (411–414). Residues 181–198 constitute a hairpin β-ribbon (PA and PAw) that wraps on the edge of P domain of another subunit. In addition to the three α-helices of residues 12–20, 393–404, and 415–429, there are a number of turns and tertiary hydrogen bonds (Supplemental Figure S3). Two tight γ-turns of Ile28-Asp30 and Asp31-Leu33 were observed in the loop connecting helix α1 and strand Sβ.

Excluding some regions that deviate significantly, the 20 models of polypeptide were refined with non-crystallographic symmetry restraint for their backbone conformations. The overall root-mean-square deviation (RMSD) is 0.085 Å between 1200 strongly restrained equivalent atoms (300 residues) in each subunit. Other equivalent backbone atoms that were also restrained deviated with RMSDs in each subunit. Other equivalent backbone atoms of Gln221 and Ser222 of subunit E also deviated by about 3 Å from the equivalent of the other subunits. These deviations reflect the flexibility of polypeptide chain in the particular regions and are well correlated with the distribution of B values (Supplemental Figure S4). Besides the terminal regions, large B values are associated with the connecting loops between β-strands, particularly the SDE, PBC, PDE, PFG, and PHP loops. In general, the internal part of the capsid shell is rigid, which comprises most of the S domain and the lower part of P domain, as shown in Fig. 2A. The B domain and the upper part of P domain show higher B values, whereas the surface loops and polypeptide termini appear to be most mobile.

In the refined model, the only residue with disallowed peptide dihedral angles of (φ, ψ) = (95°, 145°) is Ser172. This residue is located in the junction between strands SG and PA, an inter-domain region with low B values. It had to adopt such an unusual glycine-like conformation because the N atoms of Ser172 and Tyr173 formed well-defined hydrogen bonds, respectively, with the O atoms of Asp31 and Leu33 of the neighboring subunit (Supplemental Figure S3). Indeed, in other homologous virus sequences of VP2, as shown in Fig. 3A, the equivalent residue is a glycine. A cis-peptide bond was observed unambiguously between Arg347 and Pro348. This configuration occurred at the connecting segment of strand P1 to strand SH, also joining the two domains. It resulted in the “backward” hydrogen bond between the N atom of Gly344 and O atom of Arg347. Pro348 is also conserved in the birnavirus VP2 sequences. In spite of these variations, the overall subunit structure of IBDV VP2 presented here is similar to that of PDB 1WCD.

3.2. Trimer structure

It has been proposed that the VP2 trimer of IBDV is the building block for the assembly of the T = 13 virion and the T = 1 SVP because of the extensive trimer interface (Coulibaly et al., 2005). In the current structure, each subunit has a surface area of 18 510–19 020 Å², and an area of 4820–4970 Å² on each subunit is buried by its counter subunits upon trimer formation. A representative trimer is shown in Fig. 4A. The most extensive part of the interface, which is constituted by residues 169–214, covers about 1680 Å² area and encompasses the loop succeeding strand SG, the β-ribbon of PAw, and the strand Pβ. An adjacent patch of 210 Å² corresponds to the strand Pγ. These areas sum up to about 40% of the interface, and have been supposed to be an important stabilizing element of the VP2 trimer (Coulibaly et al., 2005). Another significant part of the interface (1010 Å²) corresponds to the helix α2 and its associated loops, or residues 374–412. Together with the loop of 29–35 that connects helix α1 to strand Sβ and contributes 440 Å² to the interface, this region accounts for 30% of the trimer interface. The other regions involved are residues 97–98, 128–153, 239–247, 274–300, and 331–347, located on the strands Sβ, Sβ/Sγ, Pδ, Pγ/Pβ, and P1, respectively.

Interestingly, there are ions bound at the triad axis. The calcium-binding site is 11 Å from the internal surface of the SVP. A chloride ion was also observed on the triad axis at a distance of 17 Å from the calcium ion, and 22 Å from the external surface of the SVP, as shown in Fig. 4B. The divalent cation of Ca²⁺, tightly bound to three pairs of symmetry-related Asp31 and Asp174, was observed for every trimer in the cubic crystal. All showed an identical configuration of six-coordinated octahedron, as illustrated in...
Fig. 4C. The average distances between the Ca$^{2+}$ ion and the OD1 atoms of Asp31 and Asp174 are 2.31 ± 0.09 Å and 2.19 ± 0.1 Å, respectively. On the other hand, the Cl$^-$ anion was bound to three symmetry-related peptide amide N atoms of Arg202, with a planar triangle geometry as shown in Fig. 4D, and a distance between Cl$^-$ and N of 3.15 ± 0.04 Å. The calcium-binding residues are conserved among all four birnavirus sequences (Fig. 3A), whereas the
substitution of Asp174 by an asparagine residue in *Drosophila* X virus still retains the calcium-binding capacity. Arg202 is not conserved, since its side chain is not involved.

As mentioned previously, the P domain of IBDV VP2 is similar to the rotavirus major capsid protein VP6, which also forms a trimer (Mathieu et al., 2001). The similarity extends further in that each trimer contains a metal ion bound at the threefold axis, although the type differs: Ca$^{2+}$ in IBDV and Zn$^{2+}$ in rotavirus. The zinc ion is located closer to the outer surface of the VP6 trimer, and the chloride is bound directly to the zinc. Interestingly, the IBDV VP2 trimer also contains cavities centered at the trid axis, shown in Fig. 4B. The inner cavity with an average volume of 530 Å$^3$ is located between the Ca$^{2+}$ and the Cl$^{-}$ ions, and the outer cavity of about 870 Å$^3$ between the Cl$^{-}$ and the outer surface, sealed by a single layer of Asn293

![Overall structure of the IBDV subviral particle (SVP).](image)

Fig. 2. Overall structure of the IBDV subviral particle (SVP). (A) An alpha-carbon tracing diagram of the 60-subunit SVP is viewed along an icosahedral fivefold axis. The polypeptide chains are color-coded according to the temperature factors, or $B$ values, of individual amino acid residues as observed in the cubic crystal. Deep blue denotes the coolest parts and bright red shows the hottest parts. Intermediate $B$ values are presented with ramped colors of cyan, green, yellow and orange. (B) The same object is shown but significantly truncated in both front and back to reveal the thick shell and the large central cavity. Some relevant dimensions are indicated: the tip-to-tip distance is 270 Å, the diameter of central cavity is 110 Å, and the average surface-to-surface shell diameter is 200 Å. (C) The electrostatic surface potential for a VP2 pentamer was calculated using GRASP with a range of $-15$ to $+15$ $k_B T$ ($k_B = $ Boltzmann’s constant; $T = $ Kelvin temperature). It is colored with red and blue representing negative and positive charges, and viewed along the fivefold axis of SVP.
side chains at the bottom of a bowl-shaped concave surface surrounded by three P domains. Slight alteration of the Asn293 conformation and removal of the bound Ca\(^{2+}\) and Cl\(^{-}\) ions would render a channel penetrating through the triad axis. In the T = 3 nodavirus, to which the organization of the S and B domains of IBDV VP2 bear striking resemblance, there are also bound calcium ions that stabilize the trimer structure (Wery et al., 1994), but none of them are located at the triad axis. The central channel of VP2 trimer may allow passage of other ions and participate in the viral infection process, although its precise role remains to be determined.
3.3. Roles of calcium ion

The existence of calcium ions in the SVP was verified by using ICP-MS (Supplemental Table S2). The analysis also showed that the rVP2H sample contained significant amounts of sodium and potassium, as well as trace amounts of nickel, zinc, iron, aluminum, and magnesium. The Na$^+$ and K$^+$ ions might come from the buffer and bound to the negative charged surface of VP2 (Fig. 2C). In our particle purification procedure (Lee et al., 2003), immobilized metal-ion affinity chromatography with Ni-NTA columns was used to purify the SVP particles, which may explain the occurrence of Ni$^{2+}$ ion.

As seen in Figs. 5A and B, the purified SVP showed similar mobility on agarose gel at both pH 7.0 and pH 8.5 (lane M). They remained in a particle form similar to those shown in Fig. 5C. When the samples were treated with EGTA, the bands showed up-shift on agarose gel (lane 1) but the particle integrity was maintained in both neutral and basic conditions. The EGTA treatment of rVP2H seemed to produce swollen particles that had lower mobility. After removal of the free EGTA and calcium ions, the

---

Fig. 4: Trimer structure of VP2. (A) The three triad-related subunit models are shown as ribbon diagrams in red, green and blue. The P domain of the blue model associates with that of the red via the β-ribbon of AA'. and so does the red with the green, and the green with the blue. The S domains and the B domains are also involved in the inter-subunit contacts. (B) Shown in gray are two cavities along the triad axis: an outer cavity between the outer surface of the trimer and the bound chloride ion, and an inner cavity between the chloride and the calcium ions. The protein model is colored according to the temperature factors as in Fig. 2. In the absence of both ions, these cavities would merge into a well-defined tunnel that may allow passage of other ions. (C) The calcium ion is bound to six triad-related carboxylate side chains of Asp31 and Asp174, with an octahedral configuration. (D) The chloride ion is bound to three peptide nitrogen atoms of Arg202 in a planar triangular configuration. The protein models and covalent bonds are shown in three different colors, whereas the non-covalent bonds are drawn as strings of small spheres.
resulting mobility in agarose gel was slightly restored at pH 7.0 but remained the same at pH 8.5 as in the presence of 100 mM EGTA (lane 2). It is interesting that the morphology of SVP was hardly changed at pH 7.0, but deformation and aggregation appeared at pH 8.5 (Figs. 5C and D). After the EGTA/Ca\(^{2+}\)-free rVP2H was treated with 10 mM CaCl\(_2\), the mobility of the treated sample at pH 7.0 became the same as the Ca\(^{2+}\)-bound SVP, but it still showed up-shift at pH 8.5 (lane 3).

The swelling of IBDV SVP is reversible in neutral conditions, similar to the rice yellow mottle virus (Brugidou et al., 2002). However, it is irreversible in basic conditions, even with addition of 10 mM CaCl\(_2\). The aggregates were still observed for the sample of Fig. 5D after dialysis in pH 7.0 buffer. Consequently, the swollen SVP is not stable under basic conditions, probably because the aspartate residues for binding Ca\(^{2+}\) are deprotonated and the strong repulsive force may disrupt the trimer structure. Whereas the pH dependence for T = 3 capsid structure has been observed previously (Murthy et al., 1997; Brugidou et al., 2002; Nilsson et al., 2005), similar phenomenon has not been seen before for T = 1 capsid. To the best of our knowledge, the swelling of the IBDV subviral particle at high pH is the first example for T = 1 capsid structure.

4. Discussion

4.1. Particle assembly

In the T = 13 IBDV virion or the T = 1 SVP, a VP2 trimer interacts with the other trimers mainly through the S domain, via the sides of the equilateral triangles (Coulibaly et al., 2005). In SVP, each trimer is related to adjacent trimers by the icosahedral two- and fivefold symmetry axes. The interface between trimers buries an average total surface area of 2240 Å\(^2\) with a standard deviation of 110 Å\(^2\) on each subunit (Supplemental Table S3). The additional S\(_{DE}\) loop in the models of subunit A, B, L, and T caused the average buried area to be about 230 Å\(^2\) more than the others. One subunit is in contact with three
symmetry-related subunits via the inter-trimer interface, one by the icosahedral dyad and two by the fivefold axis. Specific interactions include at least 18 direct hydrogen bonds and four non-polar contacts. Although the interface of dyad-related subunits is smaller, with an average of 380 ± 10 Å² on each subunit, it seems to be more stringent than the fivefold related interface.

A pathway for assembly of VP2 subunits into SVP is proposed in Fig. 6. The first step may involve three monomers, each having its β-ribbon arm wrapping on the counter subunit, associating into a trimer, aided by a calcium ion at the triad. Further association of the trimers is mediated through a twofold related interface, in which some stringent interactions are involved. The SVP has a large channel at each fivefold axis, sealed by the apparently disordered SDE loops. The loops were visible about the quasi-six-fold axis of the T = 13 IBDV virion, with alternating up and down conformations (Coulibaly et al., 2005). The SDE loop of PDB 1WCD has the up conformation. In our SVP structure, four subunits have visible SDE loops, three up and one down, and they are disposed about a common fivefold axis. The precursor protein pVP2, as well as RNA and other proteins including VP3, are involved in the assembly of T = 13 virion, yet recent studies showed that incorporation of an N-terminal amphipathic α-helix or retention of the C-terminal peptide segment allows VP2 to self-assemble into T = 7 or T = 13 virus-like particles (Chevalier et al., 2005; Saugar et al., 2005).

4.2. Surface charge

It is intriguing to note that purification of the SVP was achieved using immobilized metal-ion affinity chromatography, despite the fact that the appended His-tag was buried inside and not accessible to the Ni-NTA column. We analyzed the electrostatic potential of the SVP surface and found that it is generally negatively charged except the interface areas (Fig. 2C). The strong negative charge on the outer surface may account for the SVP affinity to Ni-NTA column at pH 7.8 and the particle could be eluted from the column at pH 4.0. In addition, the two residues of His249 and His253 in the PDE loop of rVP2H (Fig. 3B) could offer binding sites for the nickel ions. In fact, the heavy-atom cluster of Ta₆Br₁₂⁺ is bound to the side chains of Ser278 and Asp287 from two triad-related P subunits and His253 from the C subunit of a neighboring particle (Supplemental Figure S2).

The inner surface of SVP is also negatively charged as the outer surface (Fig. 2C). The average isoelectric point value for the disordered regions at N- (residues 1–10) and Fig. 6. Proposed assembly of IBDV SVP. The process begins with a monomer subunit (blue) associating with two other monomers (red and green). Incorporation of Ca²⁺ and Cl⁻ ions (magenta and yellow) stabilizes the trimeric structure. It continues with further association of trimers via the dyad-related interface (upper right). A possible pentamer of trimers (or pentadecamer) might be formed as an intermediate. Finally, with the addition of other trimers (gray), the pentadecamer would expand into a T = 1 subviral particle.
Asp323 made the residues 320–323 of PHI loop form a turn in the loop regions of P BC and PHI show conformational variability, and others separated by 10–20 Å, whereas those related by fivefold axis remain 40 Å apart. Capsid polymorphism as a result of different peptide lengths (Chevalier et al., 2005; Saugar et al., 2005) and variation in the inter-trimer distance may account for the slight difference in antibody cross-reactivity between the VP2 proteins with various terminal modifications (M.-Y. Wang, unpublished data). A number of other viruses display similar trimeric receptor-binding domains. Recent examples include a retrovirus envelope protein (Forster et al., 2005) and an influenza autotransporter (Yeo et al., 2004). The latter binds to receptors of the tumor necrosis factor (TNF) superfamily, which are also organized into and function as trimers (Locksley et al., 2001).

4.3. Functional loops

The three-dimensional structure of SVP enables us to correlate the hypervariable region with the four protruding loops on the particle surface, which are responsible for interaction with receptor and the recognition by antibody. The most different in sequences between serotype1 and 2 are also in the hypervariable region, especially in PDE loop, involved in virus-cell receptor binding. The two hydrophilic regions (210–225 and 312–324) on the strands P b P c and P H P i are part of the virus-neutralizing epitopes, and substitutions of the amino acids 213, 222, 318, and 323 in strain E (Fig. 3B) caused the loss of binding activity of neutralizing antibodies against strain STC (Heine et al., 1991).

In fact, these two regions were mutually stabilized by forming β-sheet structure. Because the P H loop is adjacent to P DE- neutralizing antibody binding to the epitope region may partially block the receptor-binding region (P DE and/or P FG) to interfere the viral entry. In our structure, the loop regions of P BC and PHI show conformational variations in independent VP2 subunits (Fig. 3C). However, they are stabilized by four common hydrogen bonds (Fig. 3D). The hydrogen bonds between Gln320 and Asp323 made the residues 320–323 of PHI loop form a turn conformation and interact with the P bc loop by hydrogen bonding to the backbone of Tyr220.

Heterogeneity of the surface loop conformations seems to be a result of intrinsic loop flexibility, which allows the different conformers to be captured and observed through crystal packing contacts. The two loops showing largest deviations, P bc and PHI (Fig. 3C), correspond to two segments with highest antigenic activity. In Fig. 2B, the tip-to-tip distance of an SVP is about 270 Å. The internal cavity has a diameter of about 110 Å, and the diameter of the shell is about 200 Å with a thickness of 45 Å. Consequently, the tower-like protrusion domains make the antigenic loops on the top more accessible by an additional 35 Å than other regions of the capsid surface. In the T = 1 SVP, the trimer towers are separated by an average distance of about 40 Å from each other. Such arrangement renders the VP2 protein readily recognized by antibodies or by receptors on the host cell. In the model of T = 13 virion (PDB 1WCE), some trimers are juxtaposed side by side, and others separated by 10–20 Å, whereas those related by fivefold axis remain 40 Å apart. Capsid polymorphism as a result of different peptide lengths (Chevalier et al., 2005; Saugar et al., 2005) and variation in the inter-trimer distance may account for the slight difference in antibody cross-reactivity between the VP2 proteins with various terminal modifications (M.-Y. Wang, unpublished data). A number of other viruses display similar trimeric receptor-binding domains. Recent examples include a retrovirus envelope protein (Forster et al., 2005) and an influenza autotransporter (Yeo et al., 2004). The latter binds to receptors of the tumor necrosis factor (TNF) superfamily, which are also organized into and function as trimers (Locksley et al., 2001).

4.4. Concluding remarks

The VP2 of IBDV comprises two large Swiss-roll shell (S) and protrusion (P) domains and a smaller helical base (B) domain. The similarity of the S and B domains to nodavirus and the P domain to rotavirus reveals structural relationships among the icosahedral viruses (Coulibaly et al., 2005). A similar structure in a cubic unit cell is presented here, at 2.6 Å resolution. Our data suggest that each SVP contains 20 calcium ions to stabilize the 20 trimers and regulate the swelling of the particle. Calcium ion may play important roles in IBDV assembly and infection processes, although its roles in IBDV life cycle are not clearly understood yet. The SVP showed heterogeneity in the surface loop structures. Knowledge of the three-dimensional structure may be useful in rationally incorporating important foreign epitopes into the loop region to create engineered recombinant SVP as new potent immunogens or vaccines.

5. PDB code

The structure factors of the merged data set of the cubic IBDV SVP crystal and the refined coordinates of 20 VP2 subunit models have been deposited with the Research Collaboratory for Structural Bioinformatics (RCSB) protein data bank, with Accession No. 2DF7.

Acknowledgments

The authors are grateful to Ms. Pei-Chi Chao of the Laboratory of Electron Microscopy (National Science Council) in National Chung-Hsing University, and Dr. Wen-Feng Chang of the Instrument Center at Hsinchu National Tsing-Hua University for assistance in electron microscopy and inductively coupled plasma mass spectrometry. We thank National Synchrotron Radiation Research Center of Taiwan, SPring-8 and Photon Factory of Japan for beam time allocation. We also thank Prof. Keichi Fukuyama and Prof. Tomitake Tsukihara of Osaka University, Japan, for kindly providing the Ta Br₁₃₄ compound and additional synchrotron beam time. This work was supported by grants from the National Science
Council, Taiwan (NSC-90-2317-B-005-001 and NSC-92-2317-B-005-012 to M.-Y.W., and NSC-93-3112-B-001-011-Y to A.H.-J.W. for the National Core Facility of High-Throughput Protein Crystallography at Academia Sinica).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jsb.2006.02.014.

References


Structural genomics of the Epstein–Barr virus

Epstein–Barr virus is a herpesvirus that causes infectious mononucleosis, carcinomas and immunoproliferative disease. Its genome encodes 86 proteins, which provided targets for a structural genomics project. After updating the annotation of the genome, 23 open reading frames were chosen for expression in *Escherichia coli*, initially selecting for those with known enzyme activity and then supplementing this set based on a series of predicted properties, in particular secondary structure. The major obstacle turned out to be poor expression and low solubility. Surprisingly, this could not be overcome by modifications of the constructs, changes of expression temperature or strain or renaturation. Of the eight soluble proteins, five were crystallized using robotic nanolitre-drop crystallization trials, which led to four solved structures. Although these results depended on individual treatment rather than standardized protocols, a high-throughput miniaturized crystallization screening protocol was a key component of success with these difficult proteins.

1. Introduction

Human herpesviruses comprise three subfamilies: (i) α-herpesviruses [herpes simplex viruses (HSV) 1 and 2 and varicella zoster virus (VZV)], (ii) β-herpesviruses [cytomegalovirus (CMV) and human herpesvirus (HHV) 6 and 7] and (iii) γ-herpesviruses, comprising the Kaposi’s sarcoma-associated herpesvirus (KSHV or HHV8) and Epstein–Barr virus (EBV or HHV4). The last infects the vast majority of the world’s human population, establishing and maintaining a lifelong persistence in the infected host.

Primary infection typically occurs in childhood and is frequently asymptomatic. In contrast, a delayed primary infection in adolescents or young adults results in infectious mononucleosis (IM) in approximately half of cases, with symptoms including fever, pharyngitis, lymphadenopathy and splenomegaly. IM is a self-limiting lymphoproliferative disorder characterized by an expansion of EBV-infected B-lymphocytes associated with viral lytic replication in the oropharynx, controlled by a vigorous CD8+ cytotoxic T-cell immune response. The majority of cases of acute IM recover, but serious complications can occasionally lead to death. EBV is associated with a number of cancers in the immunocompetent host (Rickinson & Kieff, 1996), in particular Burkitt’s lymphoma and nasopharyngeal carcinoma, which are endemic in African and Asian populations (Raab-Traub, 2005). Furthermore, EBV can lead to immunoproliferative disease in immunosuppressed patients, notably those infected with HIV (Rickinson & Kieff, 1996). Currently licensed antiviral drugs (acyclovir and related compounds) directed against viral DNA synthesis (Coen & Schaffer, 2003) show little effect against EBV.
Table 1
Proteins of EBV.

<table>
<thead>
<tr>
<th>Accession</th>
<th>NCBi</th>
<th>Function</th>
<th>ORF EBV</th>
<th>F</th>
<th>S</th>
<th>Occ.</th>
<th>Homologue</th>
<th>Reference</th>
<th>N</th>
<th>St</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>P03229</td>
<td>gi:140202</td>
<td>Bel-2 homologue, negative regulator of apoptosis protein BHRF1</td>
<td>BALF1</td>
<td>S</td>
<td>γ</td>
<td></td>
<td></td>
<td>Marshall et al. (1999), Bellows et al. (2002), Cabras et al. (2005)</td>
<td>1</td>
<td>†</td>
<td>Toxic</td>
</tr>
<tr>
<td>P03227</td>
<td>gi:118744</td>
<td>Single-stranded DNA-binding protein, part of replication fork/machinery</td>
<td>BALF2</td>
<td>R</td>
<td>αβγ</td>
<td>UL29</td>
<td></td>
<td>Decaussin et al. (1995), Robertson et al. (1996)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03188</td>
<td>gi:138191</td>
<td>Membrane glycoprotein B (gB, gp110) precursor, fusion and co-receptor binding</td>
<td>BALF4</td>
<td>M</td>
<td>αβγ</td>
<td>UL27</td>
<td></td>
<td>Gong et al. (1987), Spear &amp; Longnecker (2003)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03198</td>
<td>gi:118858</td>
<td>DNA polymerase</td>
<td>BALF5</td>
<td>R</td>
<td>αβγ</td>
<td>UL30</td>
<td></td>
<td>Kieff (1996), Hitt et al. (1989), Gilligan et al. (1991), Fries et al. (1997)</td>
<td>12</td>
<td>‡</td>
<td>Insoluble</td>
</tr>
<tr>
<td>Q8AZJ4</td>
<td>23893662</td>
<td>BARF0, nuclear/perinuclear in epithelial cells, start at amino acid 298 of SWISS-PROT homologue in herpes LCV tegument</td>
<td>BARF0</td>
<td>U</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03175</td>
<td>gi:132620</td>
<td>Ribonucleoside reductase, small 38 kDa subunit</td>
<td>BaRF1</td>
<td>N</td>
<td>αγ</td>
<td>UL40</td>
<td></td>
<td>Kieff (1996)</td>
<td>1</td>
<td>†</td>
<td>No expression</td>
</tr>
<tr>
<td>P03228</td>
<td>gi:115128</td>
<td>CD80 homologue (p35), oncogene, soluble glycoprotein, 2 Ig domains, binds CSF-1</td>
<td>BARF1</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td>Wei &amp; Ooka (1989), Strockbine et al. (1998), Tarbouriech et al. (2006)</td>
<td>1</td>
<td>$</td>
<td>PDB 2ch8</td>
</tr>
<tr>
<td>P03216</td>
<td>gi:136813</td>
<td>Myristoylated phosphoprotein in tegument (MyrP)</td>
<td>BBLF1</td>
<td>T</td>
<td>γ</td>
<td></td>
<td></td>
<td>Johansson et al. (2004)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not in DB</td>
<td>Not in DB</td>
<td>Primase-associated factor, spliced, full sequence not in DB, part of helicase–primase complex</td>
<td>BBLF2/3</td>
<td>R</td>
<td>αβγ</td>
<td>UL8</td>
<td></td>
<td>Yokoyama et al. (1999)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03214</td>
<td>gi:122807</td>
<td>Helicase, part of helicase–primase complex</td>
<td>BBLF4</td>
<td>R</td>
<td>αβγ</td>
<td>UL5</td>
<td></td>
<td>Yokoyama et al. (1999)</td>
<td>1</td>
<td>†</td>
<td>No expression</td>
</tr>
<tr>
<td>P03213</td>
<td>gi:136792</td>
<td>Portal protein UL6 homologue, function by homology</td>
<td>BBRF1</td>
<td>P</td>
<td>αβγ</td>
<td>UL6</td>
<td></td>
<td>Newcomb et al. (2001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P20882</td>
<td>gi:267198</td>
<td>Unknown function</td>
<td>BBRF2</td>
<td>U</td>
<td>αβγ</td>
<td>UL7</td>
<td></td>
<td>Johannsen et al. (2004), Lake &amp; Hutt-Fletcher (2000), Johannsen et al. (2004)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03215</td>
<td>gi:140278</td>
<td>Glycoprotein M, part of gN–gM complex involved in envelope–tegument interaction</td>
<td>BBRF3</td>
<td>M</td>
<td>αβγ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03226</td>
<td>gi:137569</td>
<td>Major capsid protein, MCP, VP5, 155K</td>
<td>BcLF1</td>
<td>C</td>
<td>αβγ</td>
<td></td>
<td></td>
<td>Kieff (1996)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03180</td>
<td>gi:114886</td>
<td>Viral interleukin-10 homologue precursor, VIL-10</td>
<td>BCRF1</td>
<td>S</td>
<td>βγ</td>
<td></td>
<td></td>
<td>Hsu et al. (1990), PDB 1vlk, 1y6m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q08AZK7</td>
<td>7441833</td>
<td>EBNA-Lp (EBNA-5) nuclear phosphoprotein, highly spliced, 12 exons, proline-rich, enhances EBNA-2 transactivation</td>
<td>BWRF1</td>
<td>L</td>
<td>×</td>
<td>γ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P25214</td>
<td>gi:139189</td>
<td>Minor capsid protein (mC), triplex protein HSV-1 VP23 homologue</td>
<td>BDLF1</td>
<td>C</td>
<td>αβγ</td>
<td>UL18</td>
<td></td>
<td>Trus et al. (2001)</td>
<td>1</td>
<td>†</td>
<td>Purified</td>
</tr>
<tr>
<td>Q66541</td>
<td>gi:1334910</td>
<td>Assemble without protease domain coded by 2nd reading frame</td>
<td>BDRF1</td>
<td>C</td>
<td>αβγ</td>
<td>UL26.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accession</td>
<td>NCBI</td>
<td>Function</td>
<td>ORF EBV</td>
<td>F</td>
<td>S</td>
<td>Occ.</td>
<td>Homologue</td>
<td>Reference</td>
<td>N</td>
<td>St</td>
<td>Comment</td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------</td>
<td>---</td>
<td>----</td>
<td>------</td>
<td>--------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>----</td>
<td>----</td>
<td>-----------</td>
</tr>
<tr>
<td>P03184</td>
<td>gi:136878</td>
<td>Role in DNA packaging, cytosolic zinc-binding protein, cysteine-rich</td>
<td>BRLF1</td>
<td>P</td>
<td>αβγ</td>
<td>UL32</td>
<td></td>
<td>Chang et al. (1996)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03183</td>
<td>gi:136873</td>
<td>Nuclear membrane phosphoprotein, part of intracellular virions, egress protein, complex with BFRF1</td>
<td>BRLF2</td>
<td>M</td>
<td>αβγ</td>
<td>UL31</td>
<td></td>
<td>Lake &amp; Hutt-Fletcher (2004), Gonnella et al. (2005)</td>
<td>2</td>
<td></td>
<td>Insoluble</td>
</tr>
<tr>
<td>P03185</td>
<td>gi:136886</td>
<td>Nuclear membrane protein p38, transmembrane with large cytoplasm domain, complex with BFLF2</td>
<td>BFRF1</td>
<td>M</td>
<td>βγ</td>
<td>hcmvUL50</td>
<td></td>
<td>Lake &amp; Hutt-Fletcher (2004)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14347</td>
<td>gi:140660</td>
<td>Unknown function, not included in virions</td>
<td>BFRF2</td>
<td>U</td>
<td>βγ</td>
<td>hcmvUL49</td>
<td></td>
<td>Bellows et al. (2002), Johannsen et al. (2004)</td>
<td></td>
<td>1</td>
<td>Insoluble</td>
</tr>
<tr>
<td>P14348</td>
<td>gi:139195</td>
<td>Smallest capsid protein (sCP) on outer capsid surface, HSV1-VP26 homologue, KSHV ORF 65 homologue</td>
<td>BFRF3</td>
<td>C</td>
<td>αβγ</td>
<td>UL35</td>
<td></td>
<td>Nealon et al. (2001), Bowman et al. (2003), Johannsen et al. (2004)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAA458688</td>
<td>gi:330334</td>
<td>DNA-cleavage and packaging protein, part of the DNA-packaging machinery (BFRF0.5, H54BAM)</td>
<td>BFRF4</td>
<td>P</td>
<td>αβγ</td>
<td>UL33</td>
<td></td>
<td>Alba (2002), Beard &amp; Baines (2004)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03222</td>
<td>gi:136833</td>
<td>Tegument protein (gp115)</td>
<td>BGLF1</td>
<td>T</td>
<td>γ</td>
<td></td>
<td></td>
<td>Johannsen et al. (2004)</td>
<td>1</td>
<td></td>
<td>Insoluble</td>
</tr>
<tr>
<td>P03221</td>
<td>gi:114952</td>
<td>Tegument protein, MyrPB</td>
<td>BGLF2</td>
<td>T</td>
<td>αβγ</td>
<td>UL16</td>
<td></td>
<td>Boeckmann et al. (2003), Johannsen et al. (2004)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03220</td>
<td>gi:136988</td>
<td>Homologue to HSV-1 tegument protein, but not included in virion, gp118</td>
<td>BGLF3</td>
<td>U</td>
<td>αβγ</td>
<td>UL14</td>
<td></td>
<td>Johannsen et al. (2004)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03217</td>
<td>gi:119691</td>
<td>Alkaline exonuclease, involved together with BALF2 in DNA recombination</td>
<td>BGLF5</td>
<td>R</td>
<td>αβγ</td>
<td>UL12</td>
<td></td>
<td>Cheng et al. (1980), Reuven et al. (2004)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03181</td>
<td>gi:140778</td>
<td>Proline-rich protein LF3, unknown function, tandem repeats, NotI repeat (125 bp), homologue in rhesus LCV</td>
<td>BHFL1</td>
<td>U</td>
<td></td>
<td></td>
<td></td>
<td>Laux et al. (1985), Rivailler et al. (2002), Farell (2005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03182</td>
<td>gi:119102</td>
<td>Anti-apoptotic factor bcl-2 homologue, early antigen protein R (EA-R), nuclear antigen</td>
<td>BHRF1</td>
<td>S</td>
<td>γ</td>
<td></td>
<td></td>
<td>Huang et al. (2003)</td>
<td>2</td>
<td></td>
<td>PDB 1q59 (NMR)</td>
</tr>
<tr>
<td>P03208</td>
<td>gi:138777</td>
<td>G-protein coupled receptor (G-PCR), 7 transmembrane helices, 6 glycosylation sites, 2 disulfide bridges gp94</td>
<td>BILF1</td>
<td>M</td>
<td>γ</td>
<td></td>
<td></td>
<td>Hutt-Fletcher (2005), Paulsen et al. (2005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03218</td>
<td>gi:138183</td>
<td>Membrane glycoprotein gp55/80, Ig-like, gp78</td>
<td>BILF2</td>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td>Mackett et al. (1990), Boeckmann et al. (2003), Johannsen et al. (2004)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03211</td>
<td>gi:119110</td>
<td>Latent nuclear protein EBNA-1, assures EBV episome maintenance replication, Gly-rich domain, essential for immortalization</td>
<td>BKRF1</td>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td>Bochkarev et al. (1996)</td>
<td></td>
<td></td>
<td>PDB 1b3t, 1vhi</td>
</tr>
<tr>
<td>P03212</td>
<td>gi:140976</td>
<td>Glycoprotein L precursor, gp25, in gL–gH complex involved in viral fusion together with gB</td>
<td>BKRF2</td>
<td>M</td>
<td>αβγ</td>
<td>UL1</td>
<td></td>
<td>Hutt-Fletcher (2005), Spear &amp; Longnecker (2003)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P3017</td>
<td>gi:267499</td>
<td>Tegument phosphoprotein</td>
<td>BKRF4</td>
<td>T</td>
<td>γ</td>
<td></td>
<td></td>
<td>Johannsen et al. (2004)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03200</td>
<td>gi:138368</td>
<td>Envelope glycoprotein gp350 (gp340) initial cell binding through complement receptor 2 (CR2, CD21)</td>
<td>BLLF1</td>
<td>M</td>
<td>αγ</td>
<td></td>
<td></td>
<td>Spear &amp; Longnecker (2003), Hutt-Fletcher (2005)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1 (continued)

<table>
<thead>
<tr>
<th>Accession</th>
<th>NCBI</th>
<th>Function</th>
<th>ORF EBV</th>
<th>F</th>
<th>S</th>
<th>Occ.</th>
<th>Homologue</th>
<th>Reference</th>
<th>N</th>
<th>St</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA45880</td>
<td>gi:330361</td>
<td>Envelope glycoprotein gp220, obtained through in-frame splicing from BLLF1</td>
<td>BLLF1b</td>
<td>M</td>
<td>×</td>
<td>αy</td>
<td></td>
<td>Beisel et al. (1985)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03199</td>
<td>gi:140999</td>
<td>Unknown function</td>
<td>BLLF2</td>
<td>U</td>
<td>γ</td>
<td></td>
<td>Sommer et al. (1996), Tarbouriech et al. (2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03195</td>
<td>gi:118952</td>
<td>dUTP pyrophosphatase, dUTPase</td>
<td>BLLF3</td>
<td>N</td>
<td>αy</td>
<td>U5L0</td>
<td>Lake &amp; Hutt-Fletcher (2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03196</td>
<td>gi:141001</td>
<td>Membrane glycoprotein γ1, part of the gM-gN complex, part of the envelope-temgament interaction</td>
<td>BLRF1</td>
<td>M</td>
<td>βy</td>
<td>hcmvUL73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03191</td>
<td>gi:119098</td>
<td>Processivity factor, sliding clamp, early antigen protein D (EA-D, polymerase accessory protein)</td>
<td>BMRF1</td>
<td>R</td>
<td>αfly</td>
<td>UL42</td>
<td>Tsurumi et al. (1993)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03192</td>
<td>gi:141066</td>
<td>Receptor for cellular integrins, needed for infection of epithelial cells, 10 TM helices, RGD motif</td>
<td>BMRF2</td>
<td>M</td>
<td>γ</td>
<td></td>
<td>Modrow et al. (1992), Tugizov et al. (2003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03230</td>
<td>gi:126373</td>
<td>Latent membrane protein 1 (LMP-1), interferes with signalling, TRAF-binding through CTAR1 and 2, essential for immortalization</td>
<td>BNLF1</td>
<td>L</td>
<td>αfly</td>
<td>UL37</td>
<td>Johannsen et al. (2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B28918</td>
<td>gi:4132602</td>
<td>Ribonucleoside reductase, large 140 kDa subunit</td>
<td>BORF2</td>
<td>N</td>
<td>αfly</td>
<td>UL39</td>
<td>Kieff (1996)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03186</td>
<td>gi:135574</td>
<td>Large tegument protein (LTP), viral surface protein involved in B lymphocyte binding</td>
<td>BPLF1</td>
<td>T</td>
<td>αfly</td>
<td>UL36</td>
<td>Johannsen et al. (2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P05209</td>
<td>gi:115130</td>
<td>Translation activator, R, transactivator, Rta, dimeric (TAF50)</td>
<td>BRLF1</td>
<td>S</td>
<td>γ</td>
<td></td>
<td>Hardwick et al. (1988), Graffit et al. (1990)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not in DB</td>
<td>Not in DB</td>
<td>Spliced BRLF1-BZLF1 protein, repressor of BZLF1, RAZ</td>
<td>BRLF1/ BZLF1</td>
<td>S</td>
<td>×</td>
<td></td>
<td>Manet et al. (1989), Furnari et al. (1994), Segoufin et al. (1996)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not in DB</td>
<td>Not in DB</td>
<td>Hypothetical BRLF1-BZLF1 splice variant</td>
<td>BRLF1/ BZLF1b</td>
<td>S</td>
<td>×</td>
<td></td>
<td>Farrell (2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03207</td>
<td>gi:141395</td>
<td>Enhancement of the induction of the lytic cycle</td>
<td>BRRF1</td>
<td>S</td>
<td>γ</td>
<td></td>
<td>Segoufin-Cariou et al. (2000), Hong et al. (2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03210</td>
<td>gi:141396</td>
<td>Tegument protein, unknown function</td>
<td>BRRF2</td>
<td>T</td>
<td>γ</td>
<td></td>
<td>Johannsen et al. (2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03193</td>
<td>gi:136939</td>
<td>Primase, subunit of the helicase–primase complex, part of the replication machinery</td>
<td>BSLF1</td>
<td>R</td>
<td>αfly</td>
<td>UL52</td>
<td>Yokoyama et al. (1999)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q04360</td>
<td>gi:1708410</td>
<td>mRNA-export factor (EB2, Mta,SM), mRNA splicing, interaction with human Spen proteins, IE63 (ICP27) homologue</td>
<td>BSLF2/ BMLF1</td>
<td>S</td>
<td>αfly</td>
<td>UL54</td>
<td>Boeckmann et al. (2003), Hiriart et al. (2005), Swaminathan (2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03194</td>
<td>gi:141432</td>
<td>Palmitoylated tegument protein (PalmP)</td>
<td>BSRF1</td>
<td>T</td>
<td>αfly</td>
<td>UL51</td>
<td>Johannsen et al. (2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P03119</td>
<td>gi:267575</td>
<td>Capsid maturation, capsid-associated</td>
<td>BTRF1</td>
<td>C</td>
<td>α</td>
<td>UL21</td>
<td>Wagenaar et al. (2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1 (continued)

<table>
<thead>
<tr>
<th>Accession</th>
<th>NCBI</th>
<th>Function</th>
<th>ORF EBV</th>
<th>F</th>
<th>S</th>
<th>Occ.</th>
<th>Homologue</th>
<th>Reference</th>
<th>N</th>
<th>St</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>P03235</td>
<td>gi:140616</td>
<td>EC-RF4 (ECRF4) protein</td>
<td>ECRF4</td>
<td>U</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q8AZ5J</td>
<td>gi:23893655</td>
<td>Protein LF1, contains a dUTPase like domain, γ-herpes ORF10 family</td>
<td>LF1</td>
<td>U</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q99306</td>
<td>gi:23893654</td>
<td>Protein LF2, contains a dUTPase like domain, γ-herpes ORF11 family</td>
<td>LF2</td>
<td>U</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not in DB</td>
<td>Not in DB</td>
<td>RK-BARF0, interaction with notch</td>
<td>RK-BARF0</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13285</td>
<td>gi:126379</td>
<td>LMP-2A, interference with protein kinase signalling, gene terminal protein, essential for immortalization</td>
<td>LMP-2A</td>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q8AZK9</td>
<td>gi:23893578</td>
<td>LMP-2B, negative regulator of LMP-2A</td>
<td>LMP-2B</td>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† Cloned. ‡ Expressed. § Structure solved. ¶ Purified protein. †† Crystals. ‡‡ Soluble protein.

Figure 1

(a) Classification of the EBV proteins according to function. (b) Outcome for the proteins entering into the structural genomics project.
EBV is composed of an inner capsid that contains the viral double-stranded DNA genome, surrounded by a membrane carrying various surface glycoproteins. Tegument fills the space between the capsid and the membrane. During the latent stage of infection in B-lymphocytes a very limited set of proteins is expressed. The viral DNA forms a circular episome which is associated with the cellular chromosomes and is replicated by the cellular machinery during cell division. After activation, the infection can switch to the lytic cycle, leading to the expression of the full set of viral proteins and production of viral particles. This complex lifestyle utilizes about 86 predicted proteins (Table 1), meaning that EBV has one of the largest genomes of human viruses. The principal viral functions are receptor binding and cell entry, maintenance of latency, nucleotide metabolism, DNA replication and packaging and capsid assembly (Fig. 1a, Table 1). EBV also codes for a number of immune-modulators. Some little-studied proteins shuttle viral particles from the nucleus, the site of viral replication, to the extracellular space and a number of proteins still have no assigned function. With the aim of obtaining insight into
the protein functions and in order to identify new drug targets, SPINE (Structural Genomics In Europe) included the structural proteomics of herpesviruses in workpackage 9 (human pathogen targets; see Fogg et al., 2006) and here we report our contribution to this, namely the analysis of a cohort of 23 EBV proteins.

2. Project design, methods and results

2.1. Target annotation

The project included a major continued effort in protein annotation since the information available in databases [principally SWISS-PROT (Boeckmann et al., 2003) and VIDA (Alba, 2002)] was rather incomplete, in particular for spliced reading frames, or no longer up to date. Our annotation is given in Table 1 with the results on the SPINE targets, together with as much bibliographic information as possible. We identified 86 proteins encoded by the EBV genome. The existence of a few of these remains questionable, owing to alternative splicing. The function of 15 proteins is unknown and could not be inferred from sequence homology or bibliographic information (Table 1, Fig. 1a). In general, little is known about the role of the tegument proteins, even though they have been recently localized unambiguously in the virus particle (Johannsen et al., 2004).

2.2. Target selection

As one aim of the project was to obtain structures of potential new drug targets, we first targeted proteins with known enzymatic activity (11 ORFs; Table 1). Next, proteins were ranked according to several predicted properties. Firstly, they were given priority if they had a high predicted secondary structure by the NSP@ server (Deleage et al., 1997), small size and a high stability index according to the ExPASy ProtParam tool (Gasteiger et al., 2005). Known membrane proteins, surface glycoproteins and proteins involved in the packaging mechanism were omitted in order to avoid redundancy with other teams of the SPINE project. Furthermore, we selected against components of known multi-protein assemblies and eliminated proteins containing transmembrane domains using the DAS software (Cserzo et al., 1997) and the TMHMM server (Krogh et al., 2001) available from the ExPASy web site.

2.3. Cloning and protein production

We opted for a small-scale parallel approach using simple restriction-based cloning into a vector containing a tobacco etch virus protease (TEV) cleavable N-terminal His6 tag, allowing the targets to be closely followed through purification and crystallization.

2.3.1. Cloning and expression tests. The selected genes were cloned by PCR amplification of EBV DNA extracted from the B95-8 cell line using primers introducing restriction sites at the 5’ and 3’ ends of the gene and ligated into the pPROEX-HTb plasmid (Invitrogen) using standard methods. The PCR products were cloned between NcoI or BamHI sites as a first choice, EcoRI as a second choice and HindIII or XhoI sites. The ligated products were directly transformed into Escherichia coli BL21(DE3) GOLD cells (Invitrogen), which were used for both DNA preparation for sequencing and small-scale expression tests. DNA preparation was performed either manually or automatically on the RoBioMol platform at the IBS (Grenoble). Small-scale expression tests used 1 ml LB media inoculated with single colonies. Protein production was induced with 0.5 mM isopropyl β-d-thiogalactoside and continued for 3–5 h at 310 and 303 K and overnight at 296 and 289 K. Cells were lysed with BugBuster (Novagen). Protein solubility was checked on SDS–PAGE by loading both the cell extract and the soluble fraction after centrifugation at 18 000g for 20 min. If soluble protein was not detected, the E. coli strains Rosetta, Origami, BL21 (DE3) STAR (Invitrogen), C41 and C43 (Avidis) were tested with overnight induction at 289 K.

2.3.2. Protein expression and purification. Proteins were produced using either classical LB or an auto-inducible medium (Studier, 2005). Cells were lysed by sonication and cell debris was removed by centrifugation at 30 000g for 30 min. The supernatant was loaded onto an Ni–NTA (Qiagen) column equilibrated with 20 mM Tris–HCl pH 7.5, 100 mM NaCl and 20 mM imidazole, washed using the same buffer containing 50 mM imidazole and eluted at an imidazole concentration of 500 mM. After buffer exchange back to the loading buffer, the protein was incubated overnight at room temperature with a ratio of 1/100 of recombinant His-tagged TEV protease. This was loaded again on an Ni–NTA column and the eluate of this column was concentrated by ultrafiltration and loaded onto a Superdex S75 or S200 gel-filtration column (GE/Amersham), depending on the protein size.

2.3.3. Refolding. When good expression levels of insoluble protein were obtained, refolding was attempted. Following large-scale production with induction at 310 K for 4–5 h, the protein was purified from inclusion bodies using buffers supplemented with 8 M urea. After purification and concentration to 5 mg ml−1, a 20-fold dilution in refolding buffers was followed by 24 h incubation at 277 K. Refolding buffers varied in salt concentration (0 or 500 mM NaCl), pH (Bis-Tris–HCl pH 5, Tris–HCl pH 7 or Tris–HCl pH 9) or divalent cation contents (10 mM EDTA or 5 mM CaCl2/5 mM MgCl2), leading to 12 different basic conditions. Samples were centrifuged for 15 min at 16 000g and supernatants were assayed for soluble protein either by ammonium sulfate precipitation and SDS–PAGE or by concentration followed by gel filtration.

2.4. Crystallization

Proteins were analyzed by dynamic light scattering (Protein Solutions) prior to crystallization. Crystallization screening was carried out at the High Throughput Crystallization Laboratory of the EMBL Grenoble Outstation (HTX Lab). Typically, 576 conditions were tested per sample using a PixSys4200 robot (Cartesian) and the vapour-diffusion method in CrystalQuick (Greiner Bio-One) 96-well sitting-drop crystallization plates with square wells. Drops contained 100 nl protein solution and 100 nl buffer solution. Crystal
Screen, Crystal Screen II, PEG/Ion Screen, Crystal Screen Lite, Natrix, Membrane, Grid Screens and Index Screen (Hampton Research) were used as well as Clear Strategy Screens (Molecular Dimensions). Crystallization plates were stored and automatically imaged by a CrystalMotion robot (RoboDesign) including a RoboIncubator and a Minstrel III module. Successful crystallizations were reproduced and refined manually using 1 + 1 µl hanging drops.

3. Discussion

A significant bottleneck in the structure-determination pipeline for EBV proteins was obtaining levels of protein expression (16/23) and soluble protein sufficient for crystallization (7/23; Fig. 1b, Table 1), although the success rate at crystallization was unexpectedly high (5/7). Surprisingly, changing the bacterial strain or expression temperature did not increase soluble expression levels compared with our standard protocol using BL21 cells at 30°C. A bioinformatics analysis using secondary-structure prediction (Deleage et al., 1997) and ClustalW-based alignments (Thompson et al., 1994) only rarely suggested obvious truncations. Perhaps as a consequence of this, modification of the constructs by N-terminal and C-terminal truncations, although attempted for the majority of the studied reading frames (Table 1), was successful in only one case, uracil-DNA glycosylase (UNG), where deletion of the N-terminal 24 residues increased expression levels and led to diffraction-grade crystals. The deleted residues may contain a nuclear localization signal based on sequence identity with human UNG2 (Otterlei et al., 1998). Seven soluble proteins were expressed in E. coli: the EBV protease domain, dUTPase, uracil-DNA glycosidase, BHRF1, BLRF2, BDLF1 and a fragment of BMLF1 (EB2), but the last three proteins were unstable after purification. In the case of the dUTPase, the low solubility of the protein necessitated intensive optimization of purification and crystallization conditions (Tarbourseich et al., 2005). Work on the EBV protease domain predated the SPINE project (Buisson et al., 2002). Structural determination of BHRF1 was abandoned despite the existence of small crystals when an NMR structure was reported (Huang et al., 2003). BARF1 was obtained through an external collaboration and expressed in eukaryotic cells (de Turenne-Tessier et al., 2005) before entering our structure-determination pipeline. Protein purification using an N-terminal His<sub>6</sub> tag together with a TEV protease cleavage site, sometimes including size-exclusion chromatography, reliably produced pure protein for crystallization. In line with other unpublished results in SPINE, refolding from inclusion bodies failed to produce soluble protein from any of the 12 cases. However, we subsequently tested expression in insect cells using baculovirus and obtained three soluble proteins from six ORFs. Overall in SPINE the experience has been that viral proteins tend to be more difficult to express in bacterial systems than prokaryotic proteins (e.g. 27% of viral proteins were expressed in E. coli compared with 33–77% of some bacterial proteins; Fogg et al., 2006). It is clear that eukaryotic expression is a real alternative for difficult viral proteins.

Crystallization screening used 200 nl sitting drops dispensed robotically and achieved a very high success rate; however, for proteins except BARF1 this required the addition of enzyme inhibitors (Table 2). Crystallographic details for each EBV structure are given in Table 2 and further details on the structure determinations and refinement have been or will be published elsewhere.

The study described here highlights the particular problems associated with the application of pipeline technologies to difficult proteins. In this case, EBV proteins were poorly suited to bacterial expression systems and success was dependent on a much more individual approach to protein production. Although a simple pipeline approach with standard protocols is unlikely to be universally applicable for structural determination, pipeline components can be extremely effective, exemplified here by the high-throughput nanolitre crystallization platform. This major breakthrough in crystallization screening undoubtedly contributed to the high crystallization rates observed with the soluble EBV proteins.

This work was undertaken as part of the European Union Framework Programme ‘Quality of Life and Management of Living Resources’, Integrated Project SPINE (Structural Proteomics In Europe), contract No. QLG2-CT-2002-00988. We thank Jean-Marie Seigneurin for providing DNA from the B95.8 cell line and support for the project, Florine Dupeux, Benoît Gallet, José-Antonio Marquez, Martin Rower and Thierry Vernet for the operation of high-throughput facilities at the Grenoble Partnership for Structural Biology (PSB), and Lucy Freeman and Lucie Rivail for work on individual proteins. We are grateful to Henri Gruffat and Tadamasa Ooka for help with the genome annotation.

References


Structural Basis for Specificity in the Poxvirus Topoisomerase

Kay Perry,1 Young Hwang,2 Frederic D. Bushman,2,* and Gregory D. Van Duyne1,*

1University of Pennsylvania School of Medicine
Department of Biochemistry and Biophysics and
Howard Hughes Medical Institute
242 Anatomy-Chemistry Building
2University of Pennsylvania School of Medicine
Department of Microbiology
3610 Hamilton Walk
Philadelphia, Pennsylvania 19104

Summary

Although smallpox has been eradicated from the human population, it is presently feared as a possible agent of bioterrorism. The smallpox virus codes for its own topoisomerase enzyme that differs from its cellular counterpart by requiring a specific DNA sequence for activation of catalysis. Here we present crystal structures of the smallpox virus topoisomerase enzyme bound both covalently and noncovalently to a specific DNA sequence. These structures reveal the basis for site-specific DNA recognition, and they explain how catalysis is likely activated by formation of a specific enzyme-DNA interface. Unexpectedly, the poxvirus enzyme uses a major groove binding helix that is not present in the human enzyme to recognize part of the core recognition sequence and activate the enzyme for catalysis. The topoisomerase-DNA complex structures also provide a three-dimensional framework that may facilitate the rational design of therapeutic agents to treat poxvirus infections.

Introduction

Smallpox is caused by the variola virus, a member of the Poxviridae virus family. The virus is highly transmissible with infection typically resulting in 20%–30% mortality, making it one of the most severe infectious diseases known to humans. The efficiency with which it spreads, combined with the deadly nature of the disease, has raised fears that smallpox could be revived for use in bioterrorism (Harrison et al., 2004). Structural models of smallpox virus proteins could provide the basis for rational design of antiviral agents, but few high-resolution structures of intact proteins from variola or related viruses have so far been reported (Moss, 2001).

Poxviruses are large, double-stranded DNA viruses that carry out their replication cycles entirely in the cytoplasm of infected cells. These viruses consequently encode many of the enzymes required to replicate and transcribe their genomes. Among these is a type IB topoisomerase, which is required for efficient transcription of the viral DNA (Da Fonseca and Moss, 2003). Type IB topoisomerases (TopIB) enzymes introduce transient breaks in one of the two strands of duplex DNA, allowing rotation of the flanking duplexes about the uncleaved strand (Figure 1A). These enzymes play critical roles in processes such as transcription, replication, and repair by relieving the topological stress caused by underwinding or overwinding of the DNA double helix that occurs during these events (Shuman, 1998; Wang, 1996).

The highly conserved poxvirus TopIBs are unique in several respects. They are among the smallest topoisomerases known, at only 34 kDa. Unlike the related eukaryotic cellular TopIB, which exhibits only a weak preference for certain DNA sequences (Been et al., 1984), the viral enzymes relax their substrates at specific DNA sites containing the core pentamer, 5’-(T/C)CCTT-3’ (Hwang et al., 1998; Shuman and Prescott, 1990).

Since topoisomerase activity requires the presence of the proper recognition sequence (Hwang et al., 1999a; Shuman and Prescott, 1990; Tian et al., 2004; Wittschieben and Shuman, 1997), this raises important mechanistic questions about how catalysis is coupled to sequence-specific recognition in the poxvirus enzymes. Extensive biochemical studies have been carried out to explore this issue (Koster et al., 2005; Nagarajan et al., 2005; Shuman, 1998), and structures of the isolated domains of the vaccinia virus enzyme have been reported (Cheng et al., 1998; Sharma et al., 1994). Despite a wealth of biochemical and structural data, however, progress in understanding this system has been limited by lack of structural data for the poxvirus TopIB-DNA complex. In order to establish a framework for understanding the unique features of the poxvirus topoisomerase, we have determined the crystal structures of two variola virus topoisomerase-DNA complexes, representing the noncovalent and covalent reaction intermediates shown in Figure 1A.

Results and Discussion

Complex Design and Structure Determination

We first crystallized variola TopIB (vTopIB) with a 13 bp DNA duplex containing the conserved core sequence 5’-CCCTT and optimized flanking sequences (Hwang et al., 1999a). Upon cleavage of this substrate by vTopIB (Figure 1B), the trinucleotide on the 3’ side of the cleavage site was released from the complementary strand and diffused out of the active site, trapping a covalently linked topoisomerase-DNA complex (Nunes-Duby et al., 1987). An essential step in obtaining well-diffracting crystals was the substitution of two nonconserved surface cysteine residues by serine to eliminate intermolecular disulfide bond formation. As described later, this C100S, C211S mutant is nearly as active as the wild-type enzyme in plasmid relaxation assays. The structure of the covalent vTopIB-DNA complex was determined at 2.9 Å using multil wavelength anomalous scattering from selenomethionine-substituted enzyme and then refined to a final resolution of 2.7 Å. Crystallographic data are summarized in Table 1, and representative electron density is shown in Figures 1C and 1D.

Based on these results, we designed a DNA substrate to mimic the cleavage product of the 13 bp duplex,
Figure 1. The Type IB Topoisomerase Reaction and Electron Density for the Covalent vTopIB-DNA Complex

(A) During the topoisomerase reaction cycle, the topoisomerase initially binds by wrapping around the substrate DNA with its two domains (intermediate I) (Cheng et al., 1998; Hwang et al., 1998b; Sekiguchi and Shuman, 1994; Shuman, 1998). If the correct DNA sequences are present, the enzyme is activated to carry out a transesterification reaction in which a conserved active site tyrosine residue attacks the phosphodiester on the 3' side of residue +1, resulting in the formation of a covalent 3'-phosphotyrosine linkage between the enzyme and the DNA substrate (intermediate II). This permits rotation of supercoiled DNA duplexes around the site of the nick, resulting in DNA relaxation (intermediate III) (Champoux, 2001; Koster et al., 2005; Shuman, 1998; Sivers et al., 1997). The 5'-hydroxyl group liberated during the initial cleavage event then attacks the phosphotyrosine linkage to run the cleavage reaction in reverse, resealing the DNA break, and the enzyme releases from DNA to complete the reaction cycle.

(B) Formation of the covalent topoisomerase-DNA complex. The duplex DNA substrate used for complex formation and crystallization is shown, along with the numbering scheme used throughout the text and figures. Upon cleavage of the DNA substrate, the trinucleotide on the 3' end of the cleavage site (residues −1, −2, and −3) is released, trapping the covalent enzyme-DNA complex. Note that the numbering scheme used here for poxvirus TopIB substrates differs from that used for the eukaryotic cellular topoisomerases. The enzyme attaches to the +1 nucleotide in the poxvirus TopIB convention, whereas the enzyme attaches to the −1 nucleotide in the cellular TopIB convention.

(C) Experimental electron density of the covalent vTopIB-DNA complex. The map was computed at 2.9 Å using phases from multiwavelength selenomethionine phasing and contoured at 1.2 standard deviations. The active site of the enzyme is shown, where covalent linkage between Tyr274 and the +1 phosphate is evident. Conserved catalytic residues are labelled, and important hydrogen bonds are shown.

(D) The same view as in (C), but weighted 2Fo−Fc electron density is shown following refinement at 2.7 Å.

Architecture of the Topoisomerase-DNA Complex

The covalent and noncovalent smallpox vTopIB-DNA complex structures are similar in outline, with large differences present only in the active site of the enzyme (rmsd 0.90 Å, excluding residues 264–288). The topoisomerase is folded into two domains, as anticipated from previous work on related poxviruses (Cheng et al., 1998; Cheng and Shuman, 1998; Hwang et al., 1999b; Sharma et al., 1994). The two protein domains bind on either side of the core 5'-CCCTT-3' sequence, forming a C-shaped clamp around the DNA (Figure 2A), as originally proposed based on biochemical data (Sekiguchi and Shuman, 1994). A secondary structure assignment for the full-length smallpox topoisomerase in the DNA bound structures versus those found in the isolated domains is provided in Figure S2.

The amino-terminal domain (N domain) is composed of a twisted, five-stranded antiparallel β sheet (β1–β5) with two short α helices (α1 and α2). The β5 strand of this domain is bound deeply in the major groove of the core DNA sequence, where it makes extensive direct contacts with the bases. There are very few changes in secondary or tertiary structure that occur upon DNA binding, based on comparison with the isolated N domain from vaccinia TopIB (Sharma et al., 1994). Superposition of the DNA bound variola N domain and the unbound vaccinia N domain results in an rmsd of 0.7 Å for Cα atoms.

The larger catalytic domain of vTopIB is centered on the opposite, minor groove face of the core DNA sequence, and the two domains are connected by the α3 helix. The long and sharply bent α3 helix forms the side of the C-shaped clamp (Figure 2A), passing along the DNA near positions +1 and +2, where the side chains of His76 and Arg80 contact the phosphate backbone (Figures 2A and 2B). Although the topoisomerase "clamp" formed around the core recognition site appears to be open on one face, a salt bridge between Lys65 in the [α4–]β5 hairpin and Glu139 in the α5 helix links the two domains in the noncovalent complex to fully encircle the DNA (Figure 2B). This salt bridge is not present in the covalent complex, due to an alternative choice of hydrogen bonding partners for Lys65 and Glu139. There are no significant changes in

where the scissile phosphate is present as an unyeal-1-3-terminal phosphate group. Crystallization of this substrate with vTopIB led to formation of a covalent vTopIB-DNA complex in a nearly isomorphous crystal lattice. These crystals diffracted to 1.9 Å resolution, which allowed us to clearly visualize solvent molecules in the protein/DNA interface and in the active site. The structure was refined to conventional R and Rfree values of 0.244 and 0.197, respectively (Table 1). Electron density for the noncovalent complex is shown in Figure S1 in the Supplemental Data available with this article online. Both vTopIB-DNA complex structures have been deposited with the Protein Data Bank, with accession codes 2H7F (covalent complex) and 2H7G (noncovalent complex).

DNA complex. The duplex DNA substrate used for complex formation and crystallization is shown, along with the numbering scheme used throughout the text and figures. Upon cleavage of the DNA substrate, the trinucleotide on the 3' end of the cleavage site (residues −1, −2, and −3) is released, trapping the covalent enzyme-DNA complex. Note that the numbering scheme used here for poxvirus TopIB substrates differs from that used for the eukaryotic cellular topoisomerases. The enzyme attaches to the +1 nucleotide in the poxvirus TopIB convention, whereas the enzyme attaches to the −1 nucleotide in the cellular TopIB convention.
backbone conformation in these regions, indicating that the covalent and noncovalent complexes do not differ by a domain-level opening or closing of the protein clamp around the DNA substrate.

In the human TopIB-DNA (hTopIB-DNA) complex, the protein forms a more substantially closed clamp around the DNA through interactions between loops arising from N-terminal subdomain I and the catalytic domain (Figure 2C). These interacting loops were originally referred to as the “Lips” of the topoisomerase (Redinbo et al., 1998; Stewart et al., 1998) and were more recently designated “Lip1” and “Lip2,” respectively (Patel et al., 2006). The poxvirus enzymes do not contain sequences corresponding to the Lip1 region. As shown in Figures 2B and 2C, there is also no structural equivalent of Lip2 in the vTopIB-DNA complex. In the structure of the isolated vaccinia TopIB catalytic domain (Cheng et al., 1998) and in the structure of the uncomplexed D. radiodurans TopIB (drTopIB) (Patel et al., 2006), the residues in the region corresponding to Lip2 are disordered. As shown in Figure 2B and discussed in more detail below, the corresponding region in vTopIB folds into an $\alpha$ helix when bound to DNA, and this helix plays a role in specific DNA recognition.

Overall, the vTopIB catalytic domain is primarily $\alpha$-helical ($\alpha_4$–$\alpha_{12}$) but contains a small, three-stranded $\beta$ sheet ($\beta_6$–$\beta_{18}$) that is highly conserved among the type IB topoisomerases and the tyrosine recombinases (Patel et al., 2006; Redinbo et al., 1999a; Van Duyne, 2002). A large structural reorganization of this domain occurs upon DNA binding relative to the structure of the unliganded catalytic domain (Cheng et al., 1998), with an rmsd of 3.5 Å for residues 81–310 (Figure S3 and Movie S1). The conformational change can be described as a $23^\circ$ rotation of the segment spanning helices $\alpha_4$–$\alpha_7$ and the $\beta$ sheet (Lobe1 in Figure S3), relative to the segment that includes helices $\alpha_8$–$\alpha_{12}$ (Lobe2). This subdomain rotation is crucial to formation of the enzyme active site, since the catalytic tyrosine (Tyr274) is located in Lobe2 and moves by 3.6 Å ($C_N$ atom) upon formation of the complex with DNA. The catalytic domain forms an extensive interface with the DNA substrate upstream of the cleavage site (base pairs +1 to +9), including minor groove interactions near the active site and major groove interactions involving the $\alpha_5$ helix (Figure 2A).

The DNA residues downstream of the cleavage site on the cleaved strand (positions −1 to −3) were lost in the process of trapping the covalent TopIB-DNA complex (Figure 1B); thus we cannot directly observe interactions that are present between the enzyme and the downstream sequence. However, the 5' overhang in the DNA substrate produced as a result of cleavage interacts with a symmetry-related copy of itself in the crystal lattice, taking the place of the lost trinucleotide that would normally be present 3' of the cleavage site. The enzyme makes a number of contacts with the resulting pseudocontinuous DNA duplex in this region via the

### Table 1. Summary of Crystallographic Data

<table>
<thead>
<tr>
<th></th>
<th>Noncovalent</th>
<th>Covalent (Native)</th>
<th>Covalent (SeMet MAD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>1.9</td>
<td>2.7</td>
<td>2.9</td>
</tr>
<tr>
<td>Space group</td>
<td>C222</td>
<td>C222</td>
<td>C222</td>
</tr>
<tr>
<td>Cell constants (Å)</td>
<td>a = 66.2</td>
<td>a = 66.6</td>
<td>a = 68.5</td>
</tr>
<tr>
<td></td>
<td>b = 133.7</td>
<td>b = 137.0</td>
<td>b = 137.3</td>
</tr>
<tr>
<td></td>
<td>c = 113.0</td>
<td>c = 113.2</td>
<td>c = 112.8</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.0332</td>
<td>0.97917</td>
<td>0.97952</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>96.2 (89.1)</td>
<td>99.7 (99.6)</td>
<td>97.5 (99.7)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.068 (0.532)</td>
<td>0.038 (0.285)</td>
<td>0.069 (0.305)</td>
</tr>
<tr>
<td>Total Reflections</td>
<td>422,113</td>
<td>312,043</td>
<td>311,572</td>
</tr>
<tr>
<td>Unique Reflections</td>
<td>36,403</td>
<td>14,320</td>
<td>11,747</td>
</tr>
<tr>
<td>(V/\sigma)</td>
<td>27.0 (1.57)</td>
<td>28.1 (8.23)</td>
<td>26.15 (4.88)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.2 (2.1)</td>
<td>4.3 (4.3)</td>
<td>5.8 (5.9)</td>
</tr>
</tbody>
</table>

**MAD Phasing (SOLVE)**

<table>
<thead>
<tr>
<th></th>
<th>Noncovalent</th>
<th>Covalent (Native)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z score</td>
<td>32.02</td>
<td>0.55</td>
</tr>
<tr>
<td>Number of sites found</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>3.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Noncovalent</th>
<th>Covalent (Native)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R_{merge})</td>
<td>0.243 (0.313)</td>
<td>0.237 (0.370)</td>
</tr>
<tr>
<td>(R_{work})</td>
<td>0.197 (0.273)</td>
<td>0.191 (0.315)</td>
</tr>
<tr>
<td>Number of atoms</td>
<td>2629</td>
<td>2581</td>
</tr>
<tr>
<td>Protein</td>
<td>510</td>
<td>509</td>
</tr>
<tr>
<td>DNA</td>
<td>395</td>
<td>54</td>
</tr>
<tr>
<td>Water</td>
<td>43.03</td>
<td>68.98</td>
</tr>
<tr>
<td>Average B factors (Å²)</td>
<td>43.16</td>
<td>70.16</td>
</tr>
<tr>
<td>Protein</td>
<td>49.96</td>
<td>66.36</td>
</tr>
<tr>
<td>DNA</td>
<td>0.014</td>
<td>0.011</td>
</tr>
<tr>
<td>Water</td>
<td>1.704</td>
<td>1.530</td>
</tr>
</tbody>
</table>

Numbers in parentheses represent values in highest-resolution shell.
α10a and α10b helices, strongly suggesting that the downstream DNA is contacted to at least the −2 position (data not shown). A similar conclusion was reached by modeling the DNA in the noncovalent complex as an extended DNA duplex. In the related hTopIB-DNA complex, extensive contacts are made to the downstream DNA by the coiled-coil linker (residues 636–712) and the larger N-terminal subdomains (residues 215–433), leading to a model in which these contacts control DNA rotation in the covalent intermediate (Redinbo et al., 1999a; Stewart et al., 1998). In the much smaller poxvirus enzymes, neither the coiled-coil linker nor the additional N-domain sequences are present, indicating that control of rotation in this system (Koster et al., 2005; Stivers et al., 1997) must involve a different mechanism.

Structural Basis of Sequence Specificity

The core sequence that is recognized by the poxvirus topoisomerases is 5′- (T/C)CCTT-3′, where cleavage occurs at the phosphate following the 3′-terminal thymidine (Figure 1B). In the covalent and noncovalent vTopIB/DNA crystal structures, the β5 strand in the amino-terminal domain and the α5 helix in the catalytic domain form an extensive network of major groove contacts to this core sequence (Figure 3). The side chains of residues Tyr70 and Tyr72 from β5 lie flat along the major groove, with Tyr70 covering the Cyt+3 and Cyt+4 bases and Tyr72 stacking on both the +4 ribose ring and the Thy+2 base (Figure 3A). Both tyrosine side chains also hydrogen bond to the phosphate backbone. This intimate interface explains previous observations that these residues in the vaccinia TopIB could be cross-linked to cytosines in the core DNA substrate (Sekiguchi and Shuman, 1996). A third direct contact from β5 involves Gln69, which makes a classic double hydrogen bonding interaction with Ade+2 (Figure 3A). Together, the major groove contacts involving the β5 strand explain the high degree of specificity for the +2 to +4 positions of the core recognition sequence.

It is interesting to note that subdomain I of human TopIB shares some similarity in structure with the N domain of poxvirus TopIB, including the placement of a β strand in the major groove of the DNA target (Redinbo et al., 1998). However, in the hTopIB-DNA complex, this β strand is shifted out of the groove by ~3 A relative to the position observed in the vTopIB-DNA complex, thereby preventing direct contacts to the bases. With one exception, there is little sequence similarity in this region between the poxvirus and eukaryotic cellular TopIB families. Remarkably, Tyr70 is conserved in both families of enzymes, despite playing a different role in complex formation. In the hTopIB-DNA complex, the corresponding residue (Tyr426) is both shifted out of the major groove and rotated so that it interacts only with the flanking ribose and phosphate groups (Figure 2C).

The α5 helix from the vTopIB catalytic domain also forms a complex network of contacts to the major groove of the DNA substrate (Figure 3B). In this case, water molecules play a more prominent role, forming numerous bridging hydrogen bonds that are readily visualized in the high resolution noncovalent enzyme-DNA complex (data not shown). In the core recognition sequence, Tyr136 packs against the +3 sugar and hydrogen bonds to N7 of Gua+4, while Lys133 hydrogen
bonds to both the N7 and O6 atoms of guanine in position +5. In the case of Lys133, it seems likely that a minor adjustment of the side chain would allow it to interact primarily with the N7 atom of an adenine base in the +5 position, explaining the more relaxed requirement for either Thy or Cyt on the opposite strand. Outside of the core recognition sequence, Lys135 from this helix hydrogen bonds to N7 of the +6 Gua base.

The observation of helix α5 in the vTopIB-DNA complex was not expected. This region (residues 133–143) is disordered in the structures of the vaccinia TopIB catalytic domain and the drTopIB protein. It was logical to assume that, upon binding DNA, these residues would form an ordered loop analogous to the Lip2 segment in the human TopIB/DNA structures (Figure 2C) and that this loop would interact primarily with the sugar-phosphate backbone of the DNA (Cheng et al., 1998; Patel et al., 2006). Instead, this region of poxvirus TopIB folds into an α helix and docks in the major groove where it interacts with both the bases and the backbone. The poxvirus TopIB enzyme therefore achieves its specificity for the core recognition sequence through the N domain α5 and the C domain α5 interactions with bases in the major groove. The β5 interactions specify positions +2, +3, and +4, and the α5 interactions specify positions +4 and +5.

The sequence chosen for the region upstream of the core recognition site (positions +6 to +9) in these structural studies was based on identification of an optimal target for poxvirus topoisomerases (Hwang et al., 1999a). In addition to the Lys135 interaction discussed above, vTopIB makes direct contacts to bases in this region via Arg206 and Tyr209 in the α7 helix (Figure 3C). Arg206 makes a canonical bidentate hydrogen bonding interaction with Gua+9, representing the most upstream contact between enzyme and substrate that we observe. Tyr209 makes van der Waals contact with the +6 Cyt base. As with other specific protein-DNA complexes, there are numerous polar and nonpolar interactions between the vTopIB enzyme and the sugar-phosphate backbone.

Figure 3. Specific DNA Binding and Activation of Catalysis by Smallpox Topoisomerase
(A) Specific interactions made between the β5 strand and the DNA major groove. Hydrogen bonding by Gln69, Tyr70, and Tyr72 are shown. Tyr70 and Tyr72 stack on the Cyt bases of residues +3, +4, and +5. The view is approximately the same as shown in Figure 2B.
(B) Specific interactions between the β5 helix and the DNA major groove. The view is down the β5 helical axis. The relative location of Arg130 in the active site is indicated.
(C) Interactions between α7 and the +6 to +9 residues.
(D) Close contact between the backbone amid of Gly132 and Lys133 and the +4 phosphate. This peptide segment is flanked by the β5 recognition helix and by Arg130. The view is rotated 180° relative to that shown in (B).
backbone of the DNA duplex. All of the direct vTopIB/DNA interactions observed in the noncovalent complex are summarized schematically in Figure 4A. In contrast, the hTopIB enzyme makes multiple independent contacts (e.g., Tyr70 and Tyr72) in the minor groove, an interaction that is similar to that seen in hTopIB-DNA complexes (Champoux, 2001) and in the tyrosine recombinases (Van Duyne, 2002). On the major groove face of the same +1 base pair, Arg80 from the α3 helix stacks its aromatic guanido group on the C5-methyl groups of Thy+1 and Thy+2. Together, the Lys167 and Arg80 interactions may explain the preference for Thy in the +1 position.

The availability of several human TopIB-DNA complex structures (Redinbo et al., 1998, 1999b, 2000; Stewart et al., 1998) allows us to compare the protein DNA interfaces formed by the highly specific viral TopIB to the less-specific human enzyme. In vTopIB, there are nine residues that make direct interactions with DNA bases in the major groove (Figure S4). Some side chains make multiple independent contacts (e.g., Tyr70 and Tyr72; Figure 3A). In contrast, the hTopIB enzyme makes no direct contacts to bases in the major groove, and the binding interface is almost entirely formed between the protein and the DNA backbone. Despite the differences in specificity and sizes between the two enzymes (hTopIB is 91 kDa), the amount of solvent accessible surface is remarkably similar. The vTopIB-DNA complex buries ~3100 Å² of accessible surface in the region upstream of the cleavage site (base pairs +1 to +10). The human complex buries ~2700 Å² of accessible surface in the same region. In both cases, the topoisomerase proteins contact the DNA substrate downstream of the cleavage site as well, leading to a total buried surface of ~4500 Å² in the human TopIB/DNA interface. The corresponding interface in the vTopIB complex with an extended DNA duplex is expected to be somewhat less than this, given that the viral enzyme lacks many of the protein motifs that make downstream contacts in the human enzyme.

Structural Analysis of Poxvirus TopIB Mutants

A wealth of mutagenesis data exists for the vaccinia virus TopIB enzyme that can now be interpreted in the context of the specific interface observed in the variola TopIB-DNA complex. A partial list of reported mutations (190 mutants; 146 residues) and their effects on catalysis is given in Table S1, with corresponding literature references. We have divided these mutants into two groups: those that reduce plasmid relaxation activity by 50% or more and those that do not. In Figure 4A, these mutants are mapped onto a color-coded surface of vTopIB. Most strikingly, substitutions that have the strongest effect on topoisomerase catalysis (shown in red) map almost entirely to three distinct locations: the β5 region, the α5 region, and the active site. Since mutagenesis data are not available to aid in the interpretation of some of the protein-DNA contacts observed in the vTopIB-DNA complex, we constructed an additional set of mutants and analyzed their ability to relax negatively supercoiled DNA (Figures 4B and 4C). Also included in Figure 4B are relaxation data for vTopIB active site mutants, which show the expected levels of catalytic impairment relative to those determined for the vaccinia and human TopIB enzymes (the active site residues are discussed in more detail below).

In the β5 region of vTopIB (Figure 3A), mutations of Tyr70 or Tyr72 have already been shown to result in defects in DNA binding, cleavage, and relaxation (Table S1). We analyzed the Gin69Ala mutant and found that it is also defective in relaxation (Figure 4B). Thus, mutation of any of the three residues that make direct base contacts in the vTopIB β5-DNA interface leads to defects in relaxation activity. Interestingly, all three of these residues are conserved in drTopIB, suggesting that this subbacterial TopIB enzyme may share some of the core sequence preferences identified for the viral enzymes.

In the α5 region (Figure 3B), mutation of Tyr136 to Asp or Ala caused a 100-fold drop in relaxation activity, whereas mutation of the same residue to Ser resulted in wild-type activity (Table S1). Simple modeling of the Tyr136Ser mutation in the vTopIB-DNA complex suggests that Ser would be ideally positioned to hydrogen bond to the phosphate backbone, perhaps explaining why this substitution is tolerated. The Tyr136Ala mutant was found to be more defective in the cleavage step of the reaction than in ligation, leading to the conclusion that this residue may be involved in an activation step prior to cleavage (Wittschieben and Shuman, 1997). In the context of the current structure, it seems likely that Tyr136 contributes to sequence-specific activation of catalysis rather than to closure of a clamp involving a Lip-like region, as has been suggested (Patel et al., 2006). Lys133 and Lys135 in the α5 helix also make direct contacts to bases, and their mutations to alanine result in modest 3-fold and 2-fold decreases in relaxation, respectively (Figure 4B). Neither is conserved in drTopIB, and, although Lys135 is conserved in the cellular eukaryotic enzymes, it interacts with a phosphate group in the hTopIB-DNA complex.

A particularly interesting site of previous mutagenesis in this region is Leu137, which is positioned in the middle of the α5 helix. This side chain is located on the opposite face of α5 that interacts with the DNA substrate’s major groove. Most residues would likely be accommodated as substitutions at this position, based on inspection of the structure. One residue that would not be expected to be tolerated in this position is proline, which would severely disrupt or distort the local α5 helical structure. Indeed, the Leu137His mutant has wild-type relaxation activity, but the Leu137Pro mutant is severely defective in relaxation, cleavage, and ligation (Wittschieben and Shuman, 1994).
catalysis in the TopIB family of enzymes comes from a recently described \textit{Leishmania donovani} TopIB (ldTopIB)-DNA-vanadate complex (Davies et al., 2006). Since vanadium can form a pentacoordinate complex with oxygen ligands and substitute for the normal phosphodiester linkage, this structure effectively mimics the expected transition state of the topoisomerase cleavage and ligation reactions. Consistent with the vanadate

Figure 5. Active Site of the Noncovalent \textit{variola} TopIB-DNA Complex

(A) View showing the conserved catalytic residues Arg130, Lys167, Arg223, His265, and Tyr274 as seen in the noncovalent complex. (B) View of the active site showing the hydrogen bonding network involving Glu124. Aside from the water-mediated interaction shown, this residue is buried in the hydrophobic core of the catalytic domain. (C) View of the active site showing the position of Asp188 relative to Lys167 and Arg130.
complex structure and with a great deal of structural and biochemical studies (Champoux, 2001; Nagarajan et al., 2005; Shuman, 1998), the prevailing model for phospho-
yl transfer catalysis by this family of enzymes can be described as follows: (1) the active site is assembled (if not already preactivated) by formation of the appropri-
ate enzyme-substrate complex; (2) Arg130, Arg223, and His265 stabilize the buildup of negative charge in the transition state as Tyr274 attacks the scissile phos-
phate; (3) Arg130 and Lys167 together contribute to pro-
tonation of the O5' hydroxyl leaving group (the direct proton donor is not known); and (4) following strand ro-
tation, the covalent 3'-phosphotyrosine intermediate formed upon expulsion of O5' is then the target for ligation, where the cleavage reaction is run in reverse with O5' as nucleophile and Tyr274 as the leaving group. Interestingly, no protein residue has been identi-
fied that acts as general base/acid catalyst on the Tyr274 hydroxyl group during cleavage and ligation. Instead, an active site water molecule has been sug-

The vTopIB-DNA complexes described here were de-
signed to provide high-resolution structural models of the specific enzyme-DNA interface. Both complexes lack the 5'-hydroxyl leaving group, which means that only limited new insight can be provided with respect to general acid catalysis, relative to existing structural models of TopIB and tyrosine recombinase systems. However, the covalent and noncovalent vTopIB-DNA complex structures have provided several interesting and potentially important observations with respect to the enzyme active site.

First, the most significant difference between the co-
valent and noncovalent topoisomerase-DNA complex structures is the positioning of the Tyr274 nucleophile (compare Figures 1C and 5A). The distance between the Tyr274 hydroxyl group and the scissile phosphate in the noncovalent complex is ~8 Å, indicating that a rather large movement of the x10a–x10b segment (Fig-
ure 2A and Figure S2) must occur during cleavage. This arrangement is quite different than that seen in compar-
ing the covalent versus noncovalent hTopIB-DNA com-
plexes, where the catalytic tyrosine position moves by much less (Redinbo et al., 1998).

Some of the difference in tyrosine positioning ob-
served for the viral system may be due to the lack of downstream duplex in the vTopIB-DNA complexes, which may allow the catalytic subdomain containing Tyr274 to adopt a slightly altered position. This implies that this subdomain (residues 218–314; Lobe2 in Figure S3) must be inherently quite flexible, which is consistent with the large conformational differences ob-
served between the unbound vaccinia TopIB catalytic domain (Cheng et al., 1998) and the same domain in vTopIB when bound to DNA (Figure S3 and Movie S1). Interestingly, the position of Tyr274 observed in the non-
covalent vTopIB-DNA complex is intermediate between that observed in the unliganded catalytic domain and that observed in the covalent complex with DNA, sug-

gesting that the noncovalent complex may represent a snapshot on the active site assembly pathway. The ob-
erved flexibility in the vTopIB complex is also consis-
tent with the high degree of plasticity observed when comparing multiple structures of hTopIB-DNA com-
plexes (Redinbo et al., 1999b).

A second surprising observation in the vTopIB-DNA complex active site is the unusual structural role of Glu124. This residue is almost entirely buried in the hy-
drophobic core of the catalytic domain, where one of its carboxyl oxygen atoms receives a hydrogen bond from a tightly bound water molecule located in the active site pocket (Figure 5B). The second carboxyl oxygen of this residue is 2.6 Å from the carbonyl oxygen of Ile129. This close contact requires that either the Glu124 side chain is in the neutral, protonated form or that the Ile129-Arg130 peptide bond adopts the normally disfa-
vored imidic acid tautomer. The crystallization condi-
tions used for the covalent and noncovalent topoiso-
merase-DNA complexes (pH 8) would not be expected to artificially protonate Glu124, making it unlikely that we are observing an artifact of crystallization conditions. Glu124 is conserved among the orthopox virus topoiso-
merases but is not conserved among the other eukary-
otic type IB enzymes. Mutation of this residue to Ala or Glu has little effect on enzyme relaxation (Figure 4B), in-
dicating that it does not play a crucial catalytic role. Fur-
ther work will be required to determine why the tyros-
rines have maintained this unusual structural element.

A third observation involves Asp168, an active site resi-
due that is conserved among eukaryotic TopIB en-
zymes. This residue forms a water-mediated interaction with N3 of the +1 Ade in the minor groove and is located adjacent to Lys167, which interacts with the +1 Thy base. The position of this residue is similar in structures of hTopIB-DNA complexes (Champoux, 2001) and in the IdTopIB/DNA/vanadate transition state model (Davies et al., 2006). This residue is of interest because it is close to both Lys167 and Arg130, which have been implicated in general acid catalysis (Krogh and Shuman, 2002; Na-

garajan et al., 2005). A small pocket is formed between Asp168, Lys167, and Arg130, which would be a logical place for the 5'-hydroxyl group to reside prior to the liga-
tion step of the topoisomerase reaction cycle. Curiously, we found little biochemical data available to indicate whether Asp168 is important for catalysis. A mutant hTopIB has been isolated where both Asp533 (equiva-
 lent to poxvirus Asp168) and Asp583 were substituted by glycine (Tamura et al., 1991). This double mutant is resistant to camptothecin (discussed below) and re-
tains ~10% activity compared to the wild-type enzyme (Yanase et al., 1999).

We found that the vTopIB Asp168Ala mutation results in a 60-fold drop in relaxation activity (Figure 4B), indicat-
ing that this residue may in fact play a catalytic role. It is interesting to note that the drTopIB enzyme has histidine in this position. Histidine, like aspartic acid, could partic-
ipate directly or indirectly in acid-base catalysis. The re-
lated tyrosine recombinases do not have a conserved acidic residue or histidine in an equivalent position. In-
stead, serine and threonine tend to be highly represented adjacent to the catalytic lysine (Nunes-Duby et al., 1998).

Activation of Catalysis by Sequence-Specific DNA Recognition

A goal in our structural studies of vTopIB was to under-
stand how catalysis could be coupled to sequence-
specific DNA binding. The structures described here
provide a strong indication of how this is likely to occur. The segment immediately preceding \( x \) contains the essential active site residue Arg130. As shown in Figure 3D, specific docking of \( x \) into the major groove allows the preceding polypeptide chain to form a close interaction with the phosphate backbone at the +4/+5 phosphate on the noncleaved strand. Here, the amide hydrogens of Gly132 and Lys133 are able to straddle this phosphate and position Arg130 in the active site of the enzyme. In the absence of a core recognition sequence, the \( x \) and \( b \) major groove binding elements of the enzyme would not be able to form the intimate interface shown in Figures 2 and 3, and as a result it seems unlikely that Arg130 could be properly positioned to participate in catalysis.

The hTopIB enzyme, which has a much lower level of sequence specificity, uses an alternative mechanism to position this catalytic arginine residue. Human TopIB has a loop (Lip2; Figure 2D) in place of the poxvirus helix \( a \) that contacts only the phosphate backbone of the DNA substrate (Redinbo et al., 1998; Stewart et al., 1998). A similar set of peptide backbone amide-phosphate contacts are formed adjacent to the catalytic Arg488 (equivalent of poxvirus Arg130) in this system, but in this case nonspecific contacts in the flanking loop cooperate to build this portion of the enzyme’s active site. In poxvirus TopIB, docking of \( x \) in the adjacent major groove appears to be required for assembly of an active enzyme. This idea is consistent with earlier proposals that Tyr136 (Figure 3b) is involved in an activation step prior to cleavage (Wittschieben and Shuman, 1997).

**Antiviral Compounds that Target TopIB Enzymes**

In addition to providing insight into the unique mechanistic features of the orthopox virus topoisomerases, the structures of the vTopIB-DNA complexes described here represent an obvious target for antiviral drugs against poxvirus infections. Several classes of DNA intercalating compounds that target hTopIB are currently in use as anticancer and anti-infective agents. These drugs act as cellular poisons by binding to and trapping the covalent topoisomerase-DNA complex that forms transiently during the reaction, effectively turning the enzyme into an agent that stabilizes toxic DNA breaks in the genome. The crystal structures of covalent hTopIB-DNA complexes bound by intercalating agents from several classes demonstrate that the drugs bind between the +1 and −1 base pairs of the DNA substrate and prevent religation by distancing the scissile phosphate from the 5′-hydroxyl of the −1 base (Staker et al., 2005; Staker et al., 2002).

The pharmacological properties of hTopIB and vTopIB are in general quite different (Shuman et al., 1988). For example, the hallmark hTopIB poison camptothecin has little effect on poxvirus TopIB enzyme at moderate concentrations. The vTopIB-DNA complex structures provide a plausible explanation of why camptothecin is ineffective against the viral target. A comparison of the hTopIB/DNA/camptothecin structure to a model of the corresponding vTopIB complex is shown in Figures 6A and 6B. The most striking difference between the complexes is the extent to which the drug is encapsulated by the much larger human enzyme. The extended N-terminal subdomains in hTopIB form a large flap that covers much of the intercalation site, whereas the vTopIB enzyme has no corresponding structural elements and the modeled drug is largely solvent exposed in this region.

A closeup of the two superimposed structures (Figure 6C) also reveals that the vTopIB enzyme cannot make many of the interactions to the bound drug that are observed in the human enzyme complex. Asp533 and Arg364 form a hydrogen bond network between themselves and camptothecin in the human TopIB complex. Although Asp533 is conserved in the poxvirus enzymes...
(the corresponding vTopIB residue is Asp168, discussed previously), there is no equivalent of Arg364. Indeed, the entire Lip1 region (Figure 6C, Figures 2B and 2O) is absent in the poxvirus enzymes. This region is important for sensitivity to camptothecin, since mutations in Lip1 are known to confer resistance to the drug (Chrencik et al., 2004). The comparisons shown in Figure 6 not only suggest why camptothecin is ineffective as a viral TopIB poison, but they also illustrate a likely reason why identification of alternative agents that act via the same mechanism against poxvirus TopIB has been so difficult. The minimal nature of the viral enzyme leaves few opportunities for drug-stabilizing interactions that are presumably necessary to trap and accumulate the covalent intermediate of the reaction.

Several compounds have been recently identified that are, in fact, potent inhibitors of the vaccinia TopIB enzyme’s ability to relax negatively supercoiled DNA, some with IC<sub>50</sub> values in the nM range (Bond et al., 2006). Although these compounds do not accumulate the covalent intermediate, some appear to specifically target the enzyme-substrate complex after cleavage has occurred. In principle, it may be possible to modify such compounds that bind with high affinity to the covalent complex so that they inhibit the ligation reaction in addition to inhibiting relaxation. Since the TopIB enzyme is not strictly required for poxvirus replication (Moss, 2001), inhibition of the enzyme may therefore be possible to arrest viral infection and growth. Targeting of the covalent enzyme-DNA intermediate, as occurs with hTopIB poisons, would appear to be a better strategy. A three-dimensional structure of the inhibitors in question bound to the covalent vTopIB-DNA complex would be an ideal platform for such studies and will be a focus of further work in this area.

**Experimental Procedures**

**Topoisomerase and DNA Preparations**

We first generated an expression construct encoding the variola virus topoisomerase by site-directed mutagenesis of the vaccinia virus gene (the vaccinia WR topoisomerase differs from that of variola major by three amino acid changes: D24N, E47G, and E159K). The smallpox enzyme was later modified to reduce intermolecular disulfide formation by changing both cysteine residues (C100 and C211) to serine. Variola TopIB was overexpressed in E. coli BL21(DE3) cells and purified to homogeneity by ion exchange chromatography on SP Sepharose (Pharmacia) and Uno-S (BioRad) columns, followed by size exclusion chromatography on Sephacryl S-75 (Pharmacia). The protein was concentrated in 20 mM sodium HEPES (pH 7.5) and 400 mM NaCl and stored at 4 °C. Selenomethionine-substituted protein was overexpressed using the methionine auxotroph B834(DE3) grown in EZ Rich methionine-free medium (Teknova) and the modified proteins were purified as described above for the double cysteine mutant. To assay for relaxation activity, reaction mixtures containing (per 20 μl) 50 mM Tris-HCl (pH 8.0), 0.15 M NaCl, 5% glycerol, 1 μg pUC19 plasmid DNA and 3.1 ng of topoisomerase were incubated at 25 °C. Aliquots (20 μl) were removed at various times and quenched by the addition of a solution containing glycerol, bromophenol blue, and SDS (3% final concentration). Samples were analyzed by electrophoresis through 0.8% agarose gel in TAE buffer. After staining for 15 min in 0.5 μg/ml ethidium bromide, the gel was soaked for 30 min in water, photographed, and quantified using a Storm PhosphorImager (Molecular Dynamics).

**Supplemental Data**

Supplemental Data include four figures, one table, one movie, and Supplemental References and can be found with this article online at http://www.molecule.org/cgi/content/full/23/3/343/DC1/.

**Acknowledgments**

We thank Kushol Gupta, Peng Yuan, and beamline staff at ALS 8.2.1 and 8.2.2 for assistance with synchrotron data collection. This work was supported by NIH NIAID Middle Atlantic Regional Center of Excellence Grant U54 AI057168 (to F.D.B.). G.D.V. is an Investigator of the Howard Hughes Medical Institute. The Advanced Light Source is supported by NIH NIAID Middle Atlantic Regional Center of Excellence Grant U54 AI057168 (to F.D.B.), G.D.V. is an Investigator of the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Received: February 3, 2006
Revised: May 2, 2006
Accepted: June 12, 2006
Published: August 3, 2006

**References**


Structure of Poxvirus TopIB-DNA Complex


Accession Numbers

The covalent and noncovalent TopIB-DNA complex coordinates have been deposited in the Protein Data Bank under ID codes 2H7F and 2H7G, respectively.
Cryo-Electron Tomographic Structure of an Immunodeficiency Virus Envelope Complex In Situ

Giulia Zanetti1, John A. G. Briggs2, Kay Grünewald3, Quentin J. Sattentau4*, Stephen D. Fuller1*

1 University of Oxford, Division of Structural Biology, Wellcome Trust Centre for Human Genetics, Henry Wellcome Building for Genomic Medicine, Headington, United Kingdom, 2 Department of Chemistry and Biochemistry, Ludwig-Maximilians-Universität, Munich, Germany, 3 Department of Molecular Structural Biology, Max Planck Institut für Biochemie, Martinsried, Germany, 4 The Sir William Dunn School of Pathology, Oxford University, Oxford, United Kingdom

The envelope glycoprotein (Env) complexes of the human and simian immunodeficiency viruses (HIV and SIV, respectively) mediate viral entry and are a target for neutralizing antibodies. The receptor binding surfaces of Env are in large part sterically occluded or conformationally masked prior to receptor binding. Knowledge of the unliganded, trimeric Env structure is key for an understanding of viral entry and immune escape, and for the design of vaccines to elicit neutralizing antibodies. We have used cryo-electron tomography and averaging to obtain the structure of the SIV Env complex prior to fusion. Our result reveals novel details of Env organisation, including tight interaction between monomers in the gp41 trimer, associated with a three-lobed, membrane-distal gp120 trimer. A cavity exists at the gp41–gp120 trimer interface. Our model for the spike structure agrees with previously predicted interactions between gp41 monomers, and furthers our understanding of gp120 interactions within an intact spike.

Introduction

The envelope glycoprotein (Env) complexes of human immunodeficiency virus type-1 (HIV-1) and the related lentivirus simian immunodeficiency virus (SIV) mediate viral entry into their respective host cells. The complex is initially assembled from three copies of the precursor polyprotein gp160, which are cleaved to yield a final mature trimer of heterodimers of the gp120 and gp41 subunits (reviewed in [1]). The surface glycoprotein gp120 confers cellular tropism to the virus by binding CD4 and a co-receptor (primarily CCR5 or CXCR4), and acts as a trigger for the fusogenic activity of the transmembrane glycoprotein gp41. Gp41 associates with gp120 via noncovalent interactions, is anchored into the viral lipid envelope by a transmembrane domain (reviewed in [1]), and contains a C-terminal tail that can be naturally truncated [2].

The current model for immunodeficiency virus entry posits that CD4–gp120 interaction induces and/or stabilizes a conformation in gp120 that allows the exposure and/or creation of previously occluded co-receptor binding surfaces (reviewed in [1]). Co-receptor binding triggers further rearrangement of Env. Based in part upon conserved structural features between influenza hemagglutinin protein (HA) and HIV-1 Env, it has been proposed that co-receptor engagement leading to trimer destabilization acts as a switch for gp41 to assume an extended helical conformation and insert its N-terminus into the target cell membrane (reviewed in [3]). Subsequent refolding of gp41 into a six-helix bundle brings the viral and cellular membranes into apposition, leading to lipid mixing and pore formation.

Env is the target of neutralizing antibodies in infected or immunized hosts because of its location on the virion surface. Several crystal structures of monomeric gp120 cores, lacking the hypervariable loops and N- and C-termini and complexed with specific ligands, have informed our understanding of mechanisms of viral receptor recognition and immune evasion (e.g., [4,5]). More recently, the structures of an unliganded SIV gp120 [6] and a soluble (s)CD4-ligated, HIV-1 gp120 core containing the V3 loop [7] have been determined.

Much of Env function is directed by the protein–protein interactions within the membrane-anchored trimeric spike complex. The structure of this complex has remained unsolved due to its complexity and its instability in solution. Structural studies of intact viruses are limited by the irregular morphology [8] and the low number of surface spikes on lentiviruses [9]. Consequently, many inferences have been made with regard to Env-mediated fusion and neutralizing antibody evasion in the absence of a structure-based molecular model. Recent developments in cryo-electron tomography and image processing make it possible to study the structures of irregular membrane viruses in the native state [10–12]. We have applied these advances to the study of the membrane-anchored spike complex. We have overcome the difficulties related to the low number of spikes by: (1) the use of SIV as opposed to HIV-1, because SIV Env is more...
stable than its HIV-1 counterpart [13], and is therefore more appropriate for analysis on intact virions, and (2) using an intra-viral tail-truncated form of SIV that contains higher levels of Env than do those viruses with an intact tail. Under these conditions we have collected cryo-tomographic data, and extracted and averaged the spikes to obtain a structure of Env in its native unbound conformation. Fitting of the atomic structure of unliganded SIV Env into this complex has allowed us to propose two models for the functional Env spike. SIV provides a good working model for HIV-1, since the two viruses have a high degree of sequence similarity and exploit CD4 and CCR5 as cellular receptors. These two viruses are also believed to share antibody evasion strategies.

**Results/Discussion**

**Tomographic Reconstruction**

We collected three cryo-tomographic series containing 77 virions. Each had hundreds of spikes visible on the surface (Figure 1), in stark contrast to tomographic reconstructions of NL43 HIV-1 virions [14]. The viral cores were visible, but were not analyzed since we used AT-2 (aldrithiol-2) inactivation which may not preserve native core morphology (J. A. G. Briggs, J. B. Forsdyke, H.-G. Kräusslich, and S. D. Fuller, unpublished data). The SIVmneE11S strain used here has a naturally truncated gp41 intra-viral domain of 17 amino acids [2]. This infectious natural virus variant [15] contains a greater number of Env complexes than SIV containing the full-tail gp41, facilitating tomographic analysis.

The distribution pattern of the spikes appears to be random, with some areas of local ordering (see Figure S1). This ordering may result from a high density of spikes leading to close packing limited by the diameter of the Env complex. Viral spikes were extracted from the tomographic reconstructions, aligned to one another, and then averaged using an adaptation of the image processing protocol used by Förster et al. [11]. The alignment protocol compensates for the inherent incompleteness of electron tomographic data (the ‘‘missing wedge’’), which otherwise introduces alignment bias into the reconstruction [11] (see Materials and Methods).

A total of 2,986 spikes contributed to the final map (Figures 2 and 3, and Video S1). The resolution, based on a conservative 0.5 Fourier shell correlation threshold, is 28 Å. The spike exhibits clear 3-fold symmetry and projects approximately 120 Å from the viral membrane. It consists of a stem approximately 35 Å wide and 50 Å high, capped by

---

**Synopsis**

HIV (human immunodeficiency virus) causes AIDS (acquired immunodeficiency syndrome) that is responsible for approximately 50 million infections since its first description in 1981. Antiviral therapies have made enormous progress, but a vaccine remains essential and yet elusive. The phenotypic variability of the virus (particle size varies by 3-fold) makes a structural approach difficult. Common virus surface components must be maintained to allow attachment to and penetration of host cells for infection. Reacting to these common viral components with neutralizing antibodies would allow the immune system to respond rapidly to infection and potentially serve as a basis for a vaccine. HIV (and its close relative simian immunodeficiency virus [SIV]) avoids antibody neutralization, in part by masking these essential components with flexible structural elements such as sugars and protein domains. The structural variability of the virus forced the authors to combine over a hundred electron micrographs to visualize the structure of the individual virus particles. The authors could then computationally extract the surface components and generate their average structure. This average sheds light on mechanisms of occlusion of common viral components from the immune system. This average structure could serve as a basis for effective vaccine design.

---

**Figure 1. Three-Dimensional Reconstruction of SIV Virions from Cryo-Electron Tomography**

(A) A slice through the xy plane of a reconstructed tomogram from the –6-μm defocus dataset. The spike complexes are clearly visible on the viral surface. Some of the cores appear disrupted, reflecting AT-2 treatment. Scale bar represents 100 nm.
(B) Surface rendering of one of the virions. The membrane is represented in blue, the core in red and the spikes in orange.
(C) The same virus as in (B) viewed after removing half of the viral envelope along the dotted line to reveal the core. The density between the core and the membrane has not been rendered.

DOI: 10.1371/journal.ppat.0020083.g001
three globular densities that appear to fold over the stem in a right-hand propeller orientation. The distance between the tips of the three globular domains is 110 Å. This morphology is strikingly different from that of murine leukaemia virus (MuLV), the only other viral spike reconstructed using this approach [11]. This difference in morphology is not surprising since the surface proteins of the two complexes are unrelated in sequence or structure. MuLV SU is a roughly L-shaped 70 KDa glycoprotein in which seven N-linked glycosylation sites have been identified ([16,17], whereas SIV gp120 core adopts an ovoid structure with 23 potential glycosylation sites [6]). We repeated the alignment procedures using the MuLV spike, and the enantiomer of our final structure as starting models to rule out the possibility of reference bias. In both cases we recovered the same final structure, indicating that our choice of starting model did not bias the alignment. The dimensions of our structure are consistent with molecular modelling with the CD4-bound HIV-1 gp120 crystal structure [18] and with models based on the SIV gp120 core crystal structure [6]. Negative staining of intact virions [9] gave smaller dimensions, probably reflecting artefacts related to the staining and drying of the sample. The volume of the structure, excluding the membrane, is approximately 470 nm³ using a 2 sigma contour level.

The gp41 stem appears as a compact structure with no obvious separation between the three monomers. The membrane proximal region of gp41 is characterised by a highly conserved hydrophobic region, which is thought to mediate trimer self-assembly [19]. Our data strongly support this, since we observe a compact stalk, rather than a separation into three legs as for MuLV [11]. Furthermore, the density corresponding to the gp41 stem region is shorter and slightly wider than the post-activation coiled-coil conformation [20], in agreement with the hypothesis that a dramatic conformational change is required for gp41 to extend towards, and insert into, the target cell membrane.

The shape of the complex suggests a new model for the Env trimer organisation. The volumes are consistent with the assignment of the globular domains to gp120 and the stem to gp41. The gp120 protomers appear to fold over gp41 rather than depart radially from it, contacting each other at the top of the spike. The interaction between gp120 monomers at this contact is likely to be weak, since it is not stable in the absence of gp41 [21]. This is consistent with the need for the trimer to disassemble to allow gp41 extrusion (Figure 4). A cavity in the centre of the density, visible at a 2 sigma contour level, suggests that the gp120 trimer interface is separated from the top of the gp41 trimer (Figures 2 and 3).

**Fitting of the Unbound SIV gp120 Core into the Density**

Subtracting the total volume of a 28Å low pass–filtered 44-KDa gp41 ectodomain trimer [20] from the stem region allowed us to approximately assign the remaining density to gp120. We fitted the atomic model of the unbound SIV gp120 core domain into this density. Both the atomic structure and the available density are ellipsoidal, permitting fitting in four orientations. These fits can be assessed in the light of biophysical and biochemical considerations [18]. The CD4 binding surface should be exposed on the surface of the trimer. Carbohydrates should be exposed to solvent because glycosylation sites are almost exclusively present on the exposed surfaces of proteins, and are required on Env for evasion of neutralizing antibodies. Some non-glycosylated, highly conserved sequences are likely to be buried at the trimer interface or hidden behind variable loops. Additionally, the C1 and C5 regions responsible for the interaction with gp41 [22] should be proximal to the gp120–gp41 interface. Based upon these considerations, two fits are consistent with the observed density (Figure 3). The precision of the fit is limited by the resolution of the tomographic reconstruction, and by the lack of availability of relevant crystal structures. The crystallographic structure of

---

**Figure 2.** Reconstruction of the Env Spike, Filtered to 28 Å, and Rendered
(A) A rendered view of the complex at a 2 sigma contour level. The chirality of the distal portion is evident.
(B) As in (A), rotated 60° around the 3-fold symmetry axis.
(C) A slab through the density in the orientation represented in panel (B), revealing the cavity at the centre of the structure. The membrane is coloured in cyan. The volume corresponding to gp41 has been coloured in gray, and the remaining gp120 volume in orange.
Scale bar = 100 Å.
DOI: 10.1371/journal.ppat.0020083.g002
unbound SIV gp120 core lacks the hypervariable loops V1V2 and V3, and the N- and C-terminal domains are also truncated. The missing volume constitutes a significant portion of the protein (a total of 183 residues are absent, about a third of the native protein). The mobility of the variable regions is not known: although they are very flexible in the monomer, they might be partially stabilized in the trimer. Furthermore, the conformation of the protein in the crystal cannot be assumed to reproduce exactly the conformation in vivo. This is particularly true for a protein that is subject to major structural changes and that lacks a significant portion of its structural and functional elements. A rough estimate of the volume of the map suggests that the variable loops are in their majority not visible in the reconstruction. The extent to which sugars are represented in the electron density map is unclear. For these reasons it is not possible to distinguish between the two fits only based upon the shape of the reconstructed density. It may be possible to distinguish the models based on future studies in which gp120 protomers can be more accurately oriented by binding "landmarks," such as sCD4 or antibodies with known epitopes, to the Env spike of intact virions.

Interpretation of the Fitted Structures

The two best fits have a number of features in common. In both fitted structures, all glycans are positioned so that they could project outwards from the surface of the trimer, and the variable regions V4 and V5 are positioned on the top surface of the complex, an orientation that would facilitate their recognition by antibodies [23]. The CD4 binding surface is exposed on the outer edge of each gp120 protomer, and is orientated such that access to membrane-associated CD4 on the target cell is possible.

The stem of the V1V2 loop points outwards in both fittings, implying good exposure to solvent. There is evidence that the V1V2 loop impedes antibody access to co-receptor binding sites in monomers of gp120 [24]. A previous structural model suggested that within the trimer, this may result from an interaction between the V1V2 loop and the co-receptor binding site in adjacent monomers [6]. This cannot be excluded in either of our proposed models, although the finding that V1V2 could serve to partially mask the CD4 binding pocket to protect it from antibody recognition [24] suggests that the loop acts on the same monomer to protect conserved regions.
Figure 4. Model for HIV/SIV Receptor Engagement, Based on the First Proposed Fitting

(A) The Env complex before CD4 binding. The bridging sheet is not accessible to co-receptor and antibodies, and the V3 loop is at the trimer interface.
(B) CD4 binding causes a conformational change and an overall rotation of the gp120 molecule. The V3 loop binds selectively to co-receptor, and the bridging sheet becomes exposed and available for co-receptor binding.

The C1 and C5 regions proposed to bind gp41 are located in the truncated N and C termini of the gp120 protein. The distance of the truncated ends from the gp120–gp41 interface in both our models is consistent with the folding of the missing residues towards gp41. It is also likely that the tip of gp41 diverges outwards from the stem to bind gp120.

The two models we propose differ dramatically in the orientation of the V3 loops and in the position of the co-receptor binding sites. In the first case (Figure 3A and 3B), the V3 base points towards the trimer interface. The density in the reconstruction not occupied by the crystal structure could be attributed to the V3 loop, which would run roughly parallel to the 3-fold symmetry axis, being partially exposed in the cavities between the lobes. In this trimer model, therefore, the V3 loop would be substantially masked by packaging into the trimer axis, potentially reinforced by inter-V3 bonding. In the second model (Figure 3C and 3D), the V3 loop is oriented outwards, exposed to solvent. There is no extra density corresponding to the position of V3, meaning that V3 would be highly flexible in the CD4-unbound trimer. The extent of V3 exposure before CD4 binding is dependent on the viral origin and its adaptation to growth in tissue culture. Several studies demonstrate the inability of most V3-specific monoclonal antibodies (mAb) to bind and neutralize primary isolates of HIV-1 [25], and there is little evidence for V3 loop-specific neutralizing antibody activity against SIV, whereas V3 is a target for neutralising antibodies in tissue culture-adapted HIV-1 strains [26]. CD4–gp120 engagement is thought to elicit a conformational change in Env that increases V3 exposure. This is supported experimentally by the increased accessibility of the V3 loop in the Env trimer to antibody binding and enzymic proteolysis subsequent to CD4 engagement [27,28]. It is also consistent with the highly exposed nature of the V3 loop in the structure of the V3-containing gp120–CD4–Fab complex [7]. The increased exposure of V3 in the CD4-bound conformation could be explained by a rearrangement of the trimer in the first model, or by a conformational change involving the V3 loop in the second.

The conserved co-receptor binding site, comprising the bridging sheet [6] and associated regions, is thought to be largely inaccessible in the trimer. Experimental data support this concept, since most CD4-induced (CD4i) surface-specific mAbs cannot access their epitopes on the CD4-unligated Env trimer, as evidenced by weak or absent neutralization by these mAbs [29]. The co-receptor binding region is mostly buried at the trimer interface in our first model, and only a relatively dramatic allosteric change in each gp120 protomer, or a major shift in gp120 orientation, or both, would permit its accessibility to cell surface receptors. Its position in the second model proposed is on the outside of the lobe; in this case the V3 loop could act as a protection from antibody recognition and binding.

A trimer model has been proposed previously in which the inner domain of gp120 points towards the gp120-gp41 interface and the outer domain extends outwards [6]. Such a model is inconsistent with the orientation of the major axis of the ellipsoidal gp120 density within the density corresponding to gp120 in our structure. In this previous model, the V3 loop is exposed to solvent and the co-receptor binding site is buried at the trimer interface. Our two fittings, however, make clear that either the V3 loop is also pointing towards the trimer interface, or the co-receptor binding site is on the outside of the trimer, protected from solvent by V3.

It is likely that the Env complex exhibits some conformational flexibility. Regions such as variable loops and the co-receptor binding site might be transiently exposed and hidden. The high variability in the extent of neutralization of different viral isolates by antibodies against the V3 loop and the ability of certain viruses to bind to co-receptor without the need of prior receptor binding, might reflect the absence of a unique conformation for the assembled trimer. In principle, the map we obtained may average features of more than one conformation of the Env complex. Future work will develop three-dimensional classification tools appropriate to the analysis of structural variability in the glycoprotein spikes.

The two models proposed shed light on several important antigenic and mechanistic features of the Env trimer. First, they provide models consistent with the concept that glycans and immunodominant variable loops are positioned on exposed gp120 surfaces to damp the neutralizing antibody response. Second, they confirm that only a limited gp41 surface is exposed for antibody binding, as has been proposed previously [30]. Third, they provide two possible descriptions of the position of the V3 loop and the co-receptor binding surface in the unbound spike. Finally, by limiting the number of possible gp120 fittings to two, they exclude other possible models for the trimer.

During the review of the work presented here, another
study was published reporting the three-dimensional structure of the SIV Env complex in its CD4 unbound conformation [31] obtained using a similar electron tomography based approach. Although the height and breadth of the structures are similar, the shapes of the two structures differ, markedly in the “stem region” of the spike. In contrast with our structure, each monomer in the three-dimensional map obtained by Zhu et al. [31] is multilobed and the transmembrane gp41 glycopolypeptide divided into three separate densities. The reasons for the differences between the two structures are unclear. We consider it unlikely that the differences between the two structures reflect a difference between the two virus strains. A possibility that we cannot exclude is that the two different reconstructions represent two different conformations of the glycoprotein present on the viral surface selected during the alignment and averaging.

The two studies differ in a number of details of the image processing methodology. Most notably, we have used three distinct starting models and have treated the resolution anisotropy (caused by the “missing wedge” in tomographic data due to the restricted tilt range) during the alignment process [11]. Given the novelty of the image processing procedures used in these studies, a direct and detailed comparison of the two approaches is necessary to rule out the possibility that these methodological differences are responsible for the contrasting structures.

Conclusions
The three-dimensional reconstruction of the in situ, functional trimeric Env spike is an important step forward for two reasons.

First, visualising the three-dimensional shape of the spike gives insight into potential molecular mechanisms of viral fusion and immune evasion. It suggests the presence of a gp120 trimer interface separated from the gp120–gp41 interface. This implies that disassembly of the receptor binding protein may happen in two steps: disassembly of the gp120 trimer interface followed by separation of gp120 from gp41. The first step could be triggered by CD4 binding and be necessary for co-receptor binding. The first of the two fits that we present could be an elegant explanation of this mechanism, as represented in Figure 4. In the scheme proposed, the CD4-induced rotation of gp120 monomers would simultaneously expose both the V3 loop and the bridging sheet, allowing co-receptor engagement.

Second, the three-dimensional density provides a basis for the fitting of atomic resolution structures, and the assessment of models for trimer structure. Here, these fittings lead to two possible models for the organisation of the CD4-unbound Env trimer, which contribute to advancing the knowledge of the trimeric Env assembly and hence could contribute to the improvement of structure-based drug and vaccine design.

Future studies using defined gp120 ligands will help confirm the correct fitting.

Materials and Methods
SIVmneE11S particles released into the supernatant by infected HuT 78 cells (SIVmne/Hut 78 CL E11S) were inactivated by treatment with aldrithiol-2, a mild oxidizing agent that does not alter Env function [32], then purified by sucrose gradient centrifugation as described in [33]. Samples were mixed with BSA-adsorbed 10-nm colloidal gold and vitrihed for cryo-electron microscopy [14]. Data were collected on a Philips CM300FEG transmission electron microscope (FEI, Eindhoven, The Netherlands), equipped with a Gatan GIF 2002 postcolumn energy filter (Gatan, Pleasanton, California, United States), and images were collected with a 2K × 2K Multiscan CCD camera (Gatan). The microscope was operated at 300 kV and a final magnification of 55,000×, giving a pixel size of 0.55 nm at the specimen level. Tilt series were collected in a single step, covering a minimum angular range of 123°, with an angular increment of 3°. Defocus was measured along the tilt axis after each tilt and automatically maintained. Tilt series were collected at defocus values of −6 ± 0.5 μm to −4 ± 0.5 μm, with a total electron dose of between 50 and 70 electrons/nm².

Tilted images were aligned using the gold beads as fiducial markers to within a maximum bead-positioning error of two pixels (11 Å).

Three-dimensional reconstructions were obtained using weighted back projections. Alignment, reconstruction, extraction of sub-tomograms, and further image processing were performed using the TOM (http://www.biochem.mpg.de/tom) [34] software packages implemented in MATLAB (Mathworks, Natick, Massachusetts, United States). Visualization was performed with Amira (http://www.amiravis.com) and PyMol (http://pymol.sourceforge.net).

Sub-tomograms (360 pixels) containing individual virions were extracted for use in further processing, as described in [11]. From the −6-μm defocus sub-tomograms, 300 viral spikes were identified by eye, extracted in 63-pixel cubic boxes, aligned, and then averaged. The box size was chosen to contain part of the viral membrane and intraviral material. The averaged structure was used as a template for identification of further spikes by a global cross-correlation search of spherically masked tomograms of individual virions.

A missing-wedge function was applied to the Fourier transform of the template so that no preferential orientation of the spikes was selected due to the limits on the extent of the tilted data. The spikes corresponding to the highest 200 cross-correlation peaks for each virus were extracted and manually inspected to remove false positives such as gold beads and particles overlapping with others on neighbouring viruses. A total of 2,115 spikes (from 29 virions) were selected for the −6-μm dataset, and 4,698 (from 51 virions) for the −4-μm dataset. Initial Euler angles [11] were assigned for the −6-μm dataset by assuming the spike was oriented perpendicular to the viral surface and aligning the z axis with the vector between the centre of the virus particle and the centre of the extracted spike. Phi, the rotation around this axis, was initially randomized. Fifteen alignment iterations were carried out according to [11] using an ellipsoidal mask. An initial alignment was performed only allowing rotation around the z axis. The view along z showed clear 3-fold symmetry after five iterations (Figure S2). Ten further alignment iterations were carried out allowing rotation around all axes, applying 3-fold symmetrisation to the reference. A missing wedge was applied to the Fourier transform of the appropriately rotated references in each alignment before cross-correlation with each particle [11]. Omitting this step can lead to alignment of the particles relative to the missing wedge instead of the particle signal, introducing artefacts into the final structure. An averaged reconstruction was obtained incorporating 1,820 spikes with a cross-correlation threshold set to 80% of the mean cross-correlation value. This reconstruction was then used as a starting model for alignment of spikes from the −4-μm dataset following a similar scheme as used for the −6-μm dataset. The number of spikes incorporated into the reconstruction after each iteration was increased by gradually lowering the cross-correlation coefficient threshold. After over 60% of the dataset was included, the sampling of the angular increment was increased from 3° to 2°, and the overall angular range searched increased. Aligned semi-automatic were performed every three to five complete alignments steps to avoid preferential accumulation of noise. Iterations were continued until the Fourier shell correlation (FSC) curve and the mean cross-correlation coefficient did not improve further. The final reconstruction had a resolution of 28 Å as judged using the 0.5 FSC criterion, and was low-pass filtered to this resolution.

The molecular model of SIV unliganded gp120 was initially manually fitted into the density corresponding to gp120 and then refined using the program URO [35].
Found at DOI: 10.1371/journal.ppat.0020083.sg001 (1.8 MB TIF).

Figure S2. Three-Fold Symmetry of the Env Complex
Slices on the xy plane through the membrane distal region of the Env complex (corresponding to the gp120 trimer). Panel (A) shows the density after one iteration of alignment of the in plane rotation angle phi. Panels (B–F) show the density after two to six iterations of alignment, without imposing any symmetry. The 3-fold character of the Env complex is clear from the output of these early alignment steps.

Found at DOI: 10.1371/journal.ppat.0020083.sg002 (2.2 MB TIF).

Video S1. Structure of the SIV Env Complex on the Viral Membrane
Found at DOI: 10.1371/journal.ppat.0020083.sv001 (1.0 MB WMV).

Accession Numbers
The European Molecular Biology Laboratory (EMBL)-European Bioinformatics Institute Macromolecular Database (http://www.ebi.ac.uk/msd-srv/emsearch/index.html) accession number of the cryo-electron microscopy map of the SIV spike glycoprotein structure in situ is EMD-1216. The UniProtKB/Swiss-Prot (http://www.ebi.ac.uk/swissprot) accession number for the MULV SU glycoprotein precursor is P03385 and for the SIV envelope glycoprotein precursor gp160 is P06883. The Protein Data Bank (http://www.rcsb.org/pdb) accession number for the unliganded SIV gp120 core is PDB 2BF1.

References

The authors would like to thank Friedrich Förster for many helpful discussions and for providing the MuLV map. We are grateful to our colleagues Dr. Doryen Bubeck and Dr. Samer Sourial for critical readings of the manuscript. We thank Larry Arthur, Jeffrey Lifson, and Julian Bess for supply of AT2-inactivated SIV virus preparations.

Author contributions. GZ, JAGB, KG, QJS, and SDF conceived and designed the experiments. GZ, JAGB, and KG performed the experiments. GZ, JAGB, KG, and SDF analyzed the data. KG, QJS, and SDF contributed reagents/materials/analysis tools. GZ, JAGB, KG, QJS, and SDF wrote the paper.

Funding. This work was supported by the Wellcome Trust (H3RCVY0). SDF is a Wellcome Trust Principal Research Fellow. GZ is supported by a Wellcome Trust Structural Biology Studentship. JAGB was previously supported by a Wellcome Trust Structural Biology Studentship and a fellowship from the Alexander Humboldt Foundation. He is currently an European Molecular Biology Organization (EMBO) long-term fellow. KG is supported by the Deutsche Forschungsgemeinschaft (grants GR 19901–2 and GR 19902–1). QJS is supported by grants from the Medical Research Council (United Kingdom) and an International AIDS Vaccine Initiative/United States Agency for International Development (IAVI/USAID) award.

Competing interests. The authors have declared that no competing interests exist.

Evaluation of the safety, immunogenicity, and protective efficacy of whole inactivated simian immunodeficiency virus (SIV) vaccines with conformationally and functionally intact envelope glycoproteins. AIDS Res Hum Retroviruses 20: 772–787.


Crystal structures of Nipah and Hendra virus fusion core proteins

Zhiyong Lou1,2,*, Yanhui Xu1,2,*, Kehui Xiang1, Nan Su1, Lan Qin1, Xu Li1, George F. Gao3, Mark Bartlam1,2 and Zihe Rao1,2,4,*

1 Tsinghua-Nankai-IBP Joint Research Group for Structural Biology, Tsinghua University, Beijing, China
2 National Laboratory of Biomacromolecules, IBP, Chinese Academy of Sciences, Beijing, China
3 Center for Molecular Virology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China
4 Nankai University, Tianjin, China

Keywords
crystal structure; fusion core; Hendra virus; heptad repeat; Nipah virus

The Nipah and Hendra viruses are highly pathogenic paramyxoviruses that recently emerged from flying foxes to cause serious disease outbreaks in humans and livestock in Australia, Malaysia, Singapore and Bangladesh. Their unique genetic constitution, high virulence and wide host range set them apart from other paramyxoviruses. These characteristics have led to their classification into the new genus Henipavirus within the family Paramyxoviridae and to their designation as Biosafety Level 4 pathogens. The fusion protein, an enveloped glycoprotein essential for viral entry, belongs to the family of class I fusion proteins and is characterized by the presence of two heptad repeat (HR) regions, HR1 and HR2. These two regions associate to form a fusion-active hairpin conformation that juxtaposes the viral and cellular membranes to facilitate membrane fusion and enable subsequent viral entry. The Hendra and Nipah virus fusion core proteins were crystallized and their structures determined to 2.2 Å resolution. The Hendra and Hendra fusion core structures are six-helix bundles with three HR2 helices packed against the hydrophobic grooves on the surface of a central coiled coil formed by three parallel HR1 helices in an oblique antiparallel manner. Because of the high level of conservation in core regions, it is proposed that the Nipah and Hendra virus fusion cores can provide a model for membrane fusion in all paramyxoviruses. The relatively deep grooves on the surface of the central coiled coil represent a good target site for drug discovery strategies aimed at inhibiting viral entry by blocking hairpin formation.

Abbreviations
HA, hemagglutinin; HeV, Hendra virus; HR, heptad repeat; hRSV, human respiratory syncytial virus; MAD, multiple wavelength anomalous dispersion; NiV, Nipah virus; RSV, respiratory syncytial virus; SV5, simian virus 5 or parainfluenza virus 5.
and their ability to jump species barriers have attracted
detailed attention, as they have many of the physical
attributes to serve as potential agents of bioterrorism
[5,6,9].

Paramyxoviruses are enveloped negative-stranded
RNA viruses, forming a large family (Paramyxoviri-
dae) divided into two subfamilies with five established
RNA viruses, forming a large family (Paramyxoviridae)
[6,10,11]. Like other paramyxoviruses, NiV and HeV consist of two
surface glycoproteins on the viral surface, termed the
fusion (F) protein and glycoprotein (G protein, also
called attachment protein) [7,8,12,13]. These two glyco-
proteins are both responsible for viral fusion and entry
into host cells [12,13]. The G protein initiates viral
infection by binding to the cellular receptor (attach-
ment), whereas the F protein mediates the subsequent
virus–cell membrane fusion process [12–14]. The F
protein undergoes a series of conformational changes
in the attachment and subsequent fusion process medi-
atated by the paramyxoviruses [14–16].

The F proteins of paramyxoviruses share several fea-
tures with other viral glycoproteins responsible for
membrane fusion, including the hemagglutinin (HA)
protein in influenza virus, gp160 of HIV-1, GP of
Ebola virus and the spike protein of severe acute respira-
tory syndrome virus and other coronaviruses. These
glycoproteins play a crucial role in the conformational
changes during the virus-mediated membrane fusion
process [15–20]. They are all initially synthesized as a
single-chain precursor, termed F0 in paramyxovirus,
which is then cleaved into two subunits (F1 and F2 in
paramyxovirus) by a furin-like enzyme derived from
the host cell [7]. F1 and F2 are covalently linked by a
disulfide bond, and the complex forms a trimer on the
virus envelope. The fusion peptide at the N-terminus
of F1 is highly hydrophobic and is considered to be
responsible for direct insertion of the F protein into the
cellular lipid bilayer [15,16,22]. The highly conserved
heptad repeat (HR) regions in F1, HR1 and HR2,
seemingly act as scaffolding modules. HR1 and HR2
will interact with each other to form a so-called ‘trimer
of hairpins’, ‘six-helix bundle’ or ‘fusion core’ in the
membrane fusion process. In the fusion core structure,
three HR1 helices form a central trimeric coiled coil
surrounded by three HR2 helices in an oblique antipar-
allel manner [15,16,23]. This hairpin formation aligns
the transmembrane domain in the viral membrane clos-
ely with the fusion peptide inserted into the cellular
membrane, thus facilitating membrane fusion.

There are at least three different conformations in
the membrane fusion process in the model for the viral
fusion mechanism proposed from the gp41 structure of
HIV. The first is the native (nonfusogenic) confor-
mation in which the HR1 peptides and HR2 peptides are
inaccessible. The subsequent conformation is the pre-
hairpin intermediate, in which the HR1 peptides are
exposed with their fusion peptides inserted into the tar-
get cellular membrane. The last conformation is the
fusogenic state, in which the HR1 and HR2 peptides
come together and form a highly stable coiled coil
aligning the viral and cellular membranes in juxtaposi-
tion, facilitating membrane fusion and viral entry [24].
Introduction of exogenous soluble HR1 or HR2 into
the virus infection system will block the formation of
this hairpin structure and thus inhibit viral fusion and
subsequent infection by competing with the endog-
enous HR1–HR2 interaction and holding the F protein in
the intermediate state [16,25]. Recent studies have
shown that the HR2 peptide of NiV and HeV has
strong inhibitory activity for membrane fusion in the
in vitro cell fusion system of the viruses [26]. Our previ-
ous biochemical and biophysical studies have also
shown that the complex of HR1 and HR2 in NiV or
HeV forms a typical thermostable six-helix bundle [21].
However, detailed structures of the complex and the
interaction between HR1 and HR2 in NiV or HeV
have not been reported to date.

In this study, a single chain combining the HR1 and
HR2 peptides (termed the two-helix) was constructed
for both NiV and HeV and expressed in the Escheri-
chia coli system used previously for other paramyxo-
viruses [27–33]. We have determined the crystal
structures of the NiV two-helix and HeV two-helix to
2.2 Å resolution, thus confirming the formation of a
six-helix bundle. These structures also show the typical
characteristics of NiV and HeV F proteins as members
of the Paramyxoviridae family, and provide a struc-
tural basis to explain the inhibitory effects of HR2 on
viral fusion and formation of the fusion core structure.
The results also show that the HR2 proteins of NiV
and HeV are functionally and structurally interchange-
able, and this correlates with the sequence similarity
of the HR peptides in NiV and HeV (predicted HR1
regions are identical for NiV and HeV but there is a
two amino acid difference in the HR2 regions).

Results and Discussion

Structure determination

The HR1 and HR2 regions of the NiV and HeV F
proteins consist of residues 137–178 and residues
453–485, respectively, and were predicted by a com-
puter program called learncoil-vmf [34]. The two
peptides encompassing the N-terminal and C-terminal HRs of the NiV/HeV F protein assemble into a stable trimer of heterodimers [21]. The two-helix molecules were prepared as a single chain by linking HR1 and HR2 with a linker (Fig. 1A). The NiV two-helix forms crystals with unit cell parameters \(a = 31.7 \, \text{Å}, \ b = 31.7 \, \text{Å}, \ c = 51.3 \, \text{Å}, \ \alpha = 80.7^\circ, \ \beta = 86.3^\circ \) and \(\gamma = 65.8^\circ\), and belongs to the space group P1. The crystals contain three two-helix molecules (one stable trimer) per asymmetric unit and diffract to 2.2 Å. The solvent content is estimated to be 21% with a Matthews coefficient \((V_m)\) of 1.7 Å\(^3\)Da\(^{-1}\). The HeV selenomethionyl derivative crystals belong to space group P1 with unit cell parameters \(a = 32.0 \, \text{Å}, \ b = 32.0 \, \text{Å}, \ c = 53.9 \, \text{Å}, \ \alpha = 86.0^\circ, \ \beta = 85.8^\circ \) and \(\gamma = 68.2^\circ\), and diffraction extends to 2.2 Å. Assuming the presence of three two-helix molecules (one stable trimer) per asymmetric unit, the solvent content is estimated to be 26%, with a Matthews coefficient \((V_m)\) of 1.7 Å\(^3\)Da\(^{-1}\). 

Selected data statistics are shown in Table 1.

The HeV two-helix crystal structure was determined by multiple wavelength anomalous dispersion (MAD) from a single selenomethionyl derivative crystal. Three selenium sites were located in one asymmetric unit from Patterson maps calculated with the program CNS [35]. The model was improved by cycles of manual building and refinement using the programs o [36] and CNS [35]. The structure was subsequently refined to a final \(R\)-value of 21.3% and a free \(R\)-value of 27.4%.

The NiV two-helix crystal structure was determined by molecular replacement with the HeV two-helix structure as a search model. After rotation and translation function searches with CNS [35], the model was improved by cycles of manual building and refinement using the programs o [36] and CNS [35]. The final \(R\)-value and the free \(R\)-value for the refinement were 22.5% and 28.0%, respectively.

**Overall description of the structure**

The three-dimensional structures of NiV and HeV two-helix are very similar, with an rmsd of 1.4 Å for all \(C_\alpha\) atoms, which correlates with their high amino acid sequence identity. Unless otherwise stated, we will concentrate on the structure of the NiV two-helix in the following discussion. The fusion core of NiV has a rod-shaped structure approximately 50 Å in length and with a maximum diameter of 28 Å. The NiV two-helix complex structure is a six-helix bundle comprising a trimer of NiV two-helix molecules. The center of this bundle consists of a parallel trimeric coiled coil formed by three HR1 helices, against which three HR2 helices pack in an antiparallel manner (Fig. 2A,B).
N-terminus of HR1 and the C-terminus of HR2 are located at the same end of the six-helix bundle, placing the fusion peptide and transmembrane domains close together. A region of about 270 amino acids would be located at the other end of the six-helix bundle between HR1 and HR2 in the postfusion state of NiV F protein.

The eight amino acids in the linker and several terminal residues were disordered in the electron density map and could not be traced in any of the three molecules. In one asymmetric unit of the NiV structure, the three molecules include residues 143–176 in HR1 and 455–484 in HR2, 143–175 in HR1 and 455–484 in HR2, and 143–175 in HR1 and 458–485 in HR2, respectively. In the HeV structure, the three molecules in one asymmetric unit include residues 143–176 in HR1 and 455–484 in HR2, 143–175 in HR1 and 454–484 in HR2, and 143–177 in HR1 and 457–484 in HR2, respectively. The rmsd of the NiV two-helix and the HeV two-helix is 1.3 Å for all Cα atoms.

Residues 143–176 of HR1 fold into a nine-turn α-helix that extends over the entire length of the coiled coil. As in other naturally occurring coiled coils of the fusion core, the residues in the α and d positions of the fusion core diagram representation [24] of HR1 are predominantly hydrophobic (Fig. 1B). A sequence alignment of NiV with other representative paramyxovirus fusion proteins shows that the residues in these two HR positions are highly conserved (Fig. 1B).

Residues 455–484 of HR2 form an eight-turn amphipathic α-helix stretching the entire length of the coiled coil. Each HR2 peptide packs against the long grooves formed by the interface of the three HR1 helices, and no interaction is observed between individual HR2 helices (Fig. 2A,B). The C-terminus of HR2 ends with V484, which is aligned with N143 of HR1; N143 is also the N-terminus of the HR1 domain. The N-terminus of HR2 starts with I456, which is aligned with L175 of HR1 (Fig. 2C).

### Table 1. Data collection (A) and model refinement (B) statistics.

<table>
<thead>
<tr>
<th></th>
<th>HeV two-helix</th>
<th>NIV two-helix</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wavelength (Å)</strong></td>
<td>0.9799</td>
<td>1.5418</td>
</tr>
<tr>
<td><strong>Space group</strong></td>
<td>P1</td>
<td>P1</td>
</tr>
<tr>
<td><strong>Unit cell parameters (Å)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a = 32.3 Å, b = 32.5 Å, c = 54.0 Å, x = 87.0 Å, β = 86.3 Å, γ = 67.9 Å</td>
<td>a = 32.2 Å, b = 32.9 Å, c = 53.9 Å, x = 86.3 Å, β = 86.2 Å, γ = 68.0 Å</td>
<td>a = 31.7 Å, b = 31.7 Å, c = 51.3 Å, x = 80.7 Å, β = 86.3 Å, γ = 65.8 Å</td>
</tr>
<tr>
<td><strong>Wavelength (Å)</strong></td>
<td>0.9799</td>
<td>0.9799</td>
</tr>
<tr>
<td><strong>Resolution range (Å)</strong></td>
<td>35.0–2.2</td>
<td>35.0–2.2</td>
</tr>
<tr>
<td>(2.3–2.2)</td>
<td>(2.3–2.2)</td>
<td>(2.3–2.2)</td>
</tr>
<tr>
<td><strong>Observed reflections</strong></td>
<td>67 673</td>
<td>65 774</td>
</tr>
<tr>
<td><strong>Unique reflections</strong></td>
<td>9875</td>
<td>9788</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>99.0 (96.7)</td>
<td>99.2 (98.1)</td>
</tr>
<tr>
<td><strong>I/σ (Å)</strong></td>
<td>5.9 (4.9)</td>
<td>5.5 (4.2)</td>
</tr>
<tr>
<td><strong>R_{merge} (%)</strong></td>
<td>13.4 (37.8)</td>
<td>13.2 (36.5)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>HeV two-helix</th>
<th>NIV two-helix</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R_{work}</strong></td>
<td>21.3</td>
<td>22.5</td>
</tr>
<tr>
<td><strong>R_{free}</strong></td>
<td>27.4</td>
<td>28.0</td>
</tr>
<tr>
<td><strong>Resolution range (Å)</strong></td>
<td>35.0–2.2</td>
<td>35.0–2.2</td>
</tr>
<tr>
<td><strong>Total reflections used</strong></td>
<td>9484</td>
<td>9317</td>
</tr>
<tr>
<td><strong>No. of reflections in working set</strong></td>
<td>8980</td>
<td>8282</td>
</tr>
<tr>
<td><strong>No. of reflections in test set</strong></td>
<td>504</td>
<td>434</td>
</tr>
<tr>
<td><strong>Average B factor (Å²)</strong></td>
<td>29.8</td>
<td>31.4</td>
</tr>
<tr>
<td><strong>rmsd bonds (Å)</strong></td>
<td>0.008</td>
<td>0.011</td>
</tr>
<tr>
<td><strong>rmsd angles (°)</strong></td>
<td>0.962</td>
<td>1.153</td>
</tr>
</tbody>
</table>
Interactions between HR1 and HR2

Three HR2 helices of the NiV fusion core pack against the outside of the central coiled-coil trimer in an obliquely antiparallel manner, which suggests a common interaction mode for the other well-studied Paramyxoviridae virus fusion proteins. The HR2 helices interact with HR1 mainly through hydrophobic interactions between hydrophobic residues in HR2 and the hydrophobic grooves on the surface of the central coiled coil (Fig. 3A). The interaction region of HR1 can be divided into three parts: the upper deep groove (I144–V158), the central shallow groove (V159–T164) and the lower deep groove (A165–L172) (Fig. 3B). Residues M463, I474, L481 and V484 of HR2 are anchored in the deep groove of HR1 and make a significant contribution to the hydrophobic interactions between HR1 and HR2 (Fig. 3B). Sequence comparison between NiV/HeV and other paramyxovirus fusion proteins shows that residues contributing to the HR1–HR2 interaction (e and g positions in HR1, a and d positions in HR2) are highly conserved (Figs 1B and 3B). This pattern of sequence conservation can also be shown by a helical wheel representation of one HR1 helix and one HR2 helix [21]. Sequence comparison between NiV and SV5 fusion proteins shows that five out of nine changes (including one A to V) occur in the e and g positions of HR1, and six out of nine changes (including two L to I and one I to V) occur in HR2 at the a and d positions. In contrast, only 13 out of 15 nonconservative changes occur at the outside f, b and c positions in HR1, and three out of 21 nonconservative changes occur at positions other than a and d in HR2 (Fig. 1B).

Comparison with other fusion proteins and a fusion core model for the Paramyxoviridae family

Among paramyxovirus fusion proteins, only the SV5 and human respiratory syncytial virus (hRSV) fusion core structures have been determined to date [32,37]. The NiV fusion core structure has a similar conformation to both SV5 F1 and hRSV, and can be superim-

Fig. 2. Overall views of the fusion core structure of Nipah virus (NiV). (A) Top view of the NiV F protein fusion core structure showing the three-fold axis of the trimer. (B) Side view of the NiV F protein fusion core structure showing the six-helix bundle. (C) Interactions between the termini of HR1 and HR2. HR1 and HR2 are represented by purple and golden ribbons, respectively. The interacting residues are shown as green sticks. The residues at the N-terminus and C-terminus are labeled.
posed with an rmsd of 0.68 Å and 0.67 Å between all Cα atoms, respectively. We will focus our structural comparison on NiV F and SV5 F, as the fusion core structures of hRSV and SV5 share significant similarity. Although the NiV and SV5 fusion cores share a very similar topology, they also have some significant differences.

First, the structure of the NiV F fusion core HR1 peptide (143–176) is much shorter than its counterpart in the SV5 fusion core (122–185), although HR2 has the almost same length in NiV (455–484) and SV5 (440–477) fusion cores (Fig. 4A,B). Second, the hydrophobic grooves on the surface of the central coiled coil have some significant differences, especially in the lower deep groove (Fig. 4C,D). In the structure of the NiV F fusion core, the lower deep groove formed by T164, A165, T168, V169 and L172 is much deeper than the equivalent region of the SV5 fusion core structure, formed by A157, T158, L161, G162 and V165 (Fig. 3E). This groove is so deep that we even observe that the bottom of the grooves are connected to each other and form a connective hole in the HR1 surface. Residue M463, which occupies the d position in the HR2 region and faces the center of the trimer, anchors into this groove and greatly contributes to the stability of the fusion core complex. Residue L161 in the SV5 structure makes this groove more shallow than its counterpart, T168, in NiV due to the longer hydrophobic side chain. Sequence alignment with other Paramyxoviridae viruses also shows that NiV has the shortest hydrophobic residue in the T168 position and the longest residue in the M463 position.

Although they share many differences from other Paramyxoviridae fusion core proteins, the NiV and HeV F fusion cores also share certain similarities and show high conservation. Among the Paramyxoviridae, the NiV and HeV fusion cores have the shortest structures and sequences. However, all paramyxovirus fusion cores share the same core parts and are highly conserved, both in sequence (Fig. 1B) and in three-dimensional structure (Fig. 3B). These facts suggest that the structure of the NiV F fusion core may share
common features with all Paramyxoviridae virus fusion cores, leading us to propose the NiV F fusion core structure as a model for Paramyxoviridae fusion cores. Furthermore, the conserved deep grooves at both ends of the NiV fusion core may provide a structural basis for the design of wide-spectrum therapeutics targeting the Paramyxoviridae family.

Conformational change and membrane fusion mechanisms

Structural studies of the influenza virus HA and HIV gp41 have established a paradigm for understanding the mechanisms of viral and cellular membrane fusion [18]. The similarity between the NiV F protein and other widely studied viral fusion proteins, as well as previous biochemical analysis [38], indicates a similar mechanism of membrane fusion mediated by the NiV and HeV fusion proteins. The structures of the NiV and HeV fusion cores reported here add to the repertoire of paramyxovirus six-helix bundle fusion core structures, providing greater structural information in order to understand the formation of the fusion-active state of genus Henipavirus. Similar to SV5F and HIV gp41, the NiV and HeV fusion proteins probably undergo a series of conformational changes to become fusion-active. The
fusion loop, which inserts into the cellular membrane, is accepted to have the distinct conformational states proposed for the NiV F protein fusion core, including the native state, the prehairpin intermediate, and the fusion-active hairpin state. Several biological and inhibition studies have also provided good evidence that the fusion core in the crystal structure presented here is the final, stable form of the protein, which is the fusion-active state following one or more conformational changes. First, gel filtration and chemical crosslinking results demonstrated that the oligomeric state of the two-helix protein was a trimer. Even at high concentrations of the crosslinker, the monomer/dimer bands could be observed [31]. Second, NiV and HeV infection in vitro can be potently blocked by peptides corresponding to the C-terminal HR (HR2) of the HeV fusion envelope glycoprotein 39. These features suggest that the NiV F protein also undergoes a conformational change mechanism, similar to influenza HA and HIV gp41.

Inhibitors of NiV/HeV infection

As membrane fusion is a very important process during virus infection, inhibition studies have been carried out to find effective drugs to block virus infection by targeting the membrane fusion step. In the case of HIV-1, several strategies to block hairpin formation have been successfully developed to identify viral entry inhibitors that bind to the hydrophobic pocket and grooves on the surface of the central coiled coil consisting of HIV-1 gp41 N peptides. These useful viral entry inhibitors include D peptides, five-helix, and synthetic peptides derived from N or C peptides [40-42]. Successful viral entry inhibitors have also been identified for other viruses, such as T20 for HIV-1 and GP610 for Ebola virus. Analogous strategies could also be used for the design of NiV/HeV fusion inhibitors.

In 2005, several reports showed that NiV/HeV infection in vitro can be potently blocked by specific HR2 peptides. The improved second-generation HR2 peptides, which use poly(ethylene glycol) to facilitate peptide synthesis and increase solubility, also show good IC50 values in in vitro assays. The applied chemical modifications are also predicted to increase the serum half-life in vivo and should increase the chances of success in the development of an effective antiviral therapy [50]. The well-defined hydrophobic grooves on the surface of the central coiled coil of the NiV F protein fusion core identified from our structure can offer a reasonable explanation for the inhibition of NiV and HeV infection. Furthermore, the structures reported here provide significant targets for the design of NiV and HeV antiviral agents.

Experimental procedures

Purification and crystallization – the two-helix constructs of both NiV and HeV

Fusion proteins were prepared as a single chain by linking the HR1 and HR2 domains with an eight amino acid linker (GGSGGSSGG). The PCR-directed gene was inserted into the pET22b vector (Novagen, Shanghai, China), and the target plasmids were transformed into BL21 (DE3) competent cells. The cells were cultured at 310 K in 2 × YT medium containing 100 μg·mL⁻¹ ampicillin and were induced with 0.2 mM isopropyl thio-β-D-galactoside (IPTG) when the culture density (D₆₅₀) reached 0.6-0.8. The selenomethionine derivative HeV two-helix protein was expressed in M9 medium containing 30 mg·L⁻¹ selenomethionine in E. coli strain BL21 (DE3). The two products were both purified by nickel-nitrilotriacetic acid affinity chromatography followed by gel filtration chromatography. The purified NiV two-helix and HeV two-helix derivative were dialyzed against crystallization buffer (10 mM Tris/HCl, pH 8.0, 10 mM NaCl) and concentrated to 10–15 mg·mL⁻¹. Initial crystallization conditions were screened using Crystal Screen reagent kits I and II (Hampton Research, Aliso Viejo, CA, USA) and a poly(ethylene glycol) screening kit prepared in-house.

Good-quality NiV two-helix crystals were obtained from 0.1 M Tris/HCl (pH 8.5)/29% poly(ethylene glycol) 4000 (v/v). Good-quality HeV two-helix derivative crystals were obtained from 0.1 M Hepes (pH 6.5)/10% poly(ethylene glycol) 4000 (v/v). The preparation and crystallization of the two-helix proteins of NiV and HeV have previously been reported in detail [21].

Data collection and processing

Data collection from the NiV two-helix crystal was performed in-house on a Rigaku RU200 (Tokyo, Japan) rotating-copper-anode X-ray generator operated at 48 kV and 98 mA (CuKα; λ = 1.5418 Å) with an Mar345 image-plate detector. The crystal was mounted on nylon loops and flash-cooled in a cold nitrogen gas stream at 100 K using an Oxford Cryosystems (Oxford, UK) cold stream and with the reservoir solution as cryoprotectant. Data were indexed and scaled using the HKL2000 programs DENZO and SCALEPACK [43]. The HeV two-helix selenomethionine derivative crystal was mounted on nylon loops and flash-frozen in a cold nitrogen gas stream at 100 K using an Oxford Cryosystems cold stream and with 0.1 M Hepes (pH 6.5)/25% poly(ethylene glycol) 400 as cryoprotectant. MAD data were collected by a rotation method using a Mar CCD detector with synchrotron radiation beamline 3W1A of the Beijing Synchrotron Radiation Facility. Data were collected from a single selenomethyl derivative crystal at peak (0.9799 Å), edge (0.9801 Å) and
remote (0.9500 Å) wavelengths to 2.2 Å. Data were indexed and scaled using DENZO and SCALEPACK programs [43].

**Phase determination and model refinement**

For determination of the HeV two-helix structure, initial MAD phasing steps were performed using SOLVE [44], and density modification was performed using RESOLVE [45]. The program o [36] was used for manual tracing of the experimental density map, and the initial structure was subsequently refined using the programs o [36] and CNS [35]. The NiV two-helix structure was determined by molecular replacement with the HeV two-helix structure as a search model. Rotation and translation function searches were performed with the program CNS [35]. The model was further improved by manual building and refinement using the programs o [36] and CNS [35]. The quality of the two structures was verified by PROCHECK [46], with none of the main-chain torsion angles located in disallowed regions of the Ramachandran plot. Structure determination and refinement statistics are summarized in Table 1. The figures were generated with the programs GRASP [47], PYMOL [47] and MOLSCRIPT [48].

**Accession codes**

Coordinates and structure factors for the NiV and HeV fusion core crystal structures have been deposited in the RCSB PDB with accession numbers 1WP7 and 1WP8, respectively.

**Acknowledgements**

This work was supported by the NSFC (grant number 30221003).

**References**

The Crystal Structure of the Venezuelan Equine Encephalitis Alphavirus nsP2 Protease

Andrew T. Russo,1 Mark A. White,1 and Stanley J. Watowich1,*
1 Department of Biochemistry and Molecular Biology and Sealy Center for Structural Biology and Molecular Biophysics
The University of Texas Medical Branch
Galveston, Texas 77555

Summary

Alphavirus replication and propagation is dependent on the protease activity of the viral nsP2 protein, which cleaves the nsP1234 polyprotein replication complex into functional components. Thus, nsP2 is an attractive target for drug discovery efforts to combat highly pathogenic alphaviruses. Unfortunately, antiviral development has been hampered by a lack of structural information for the nsP2 protease. Here, we report the crystal structure of the nsP2 protease (nsP2pro) from Venezuelan equine encephalitis alphavirus determined at 2.45 Å resolution. The protease structure consists of two distinct domains. The nsP2pro N-terminal domain contains the catalytic dyad cysteine and histidine residues organized in a protein fold that differs significantly from any known cysteine protease or protein folds. The nsP2pro C-terminal domain displays structural similarity to S-adenosyl-L-methio- nine-dependent RNA methyltransferases and provides essential elements that contribute to substrate recognition and may also regulate the structure of the substrate binding cleft.

Introduction

Venezuelan equine encephalitis alphavirus (VEEV) is a significant cause of human and livestock disease in Central and South America, with outbreaks occasionally reaching as far north as southern Texas (Weaver et al., 1996). Human epidemics reported in 1995 in Venezuela and Colombia caused widespread illness with mortality rates approaching 1% (Weaver et al., 1996). Several nations, including the United States and the former Soviet Union, have reportedly developed weapons-grade VEEV (Bronze et al., 2002), and stockpiles of such weapons may still exist. Consequently, VEEV is classified as a select agent and a National Institutes of Health Category B priority pathogen. Existing health care resources are poorly prepared to deal with an outbreak of VEEV. No antiviral drugs exist to treat VEEV infection. Moreover, the VEEV TC-83 vaccine strain provides only partial protection against infection and is not approved for general human immunization (Weaver et al., 1999).

VEEV is an enveloped, positive-sense ssRNA virus representative of the family Togaviridae and genus Alphavirus (Strauss and Strauss, 1994). After infection, alphavirus RNA is directly translated to typically produce polyprotein nsP1234 containing nonstructural proteins nsP1–nsP4 (Strauss and Strauss, 1994). In VEEV, an opal codon between nsp3 and nsp4 results in the expression of polyprotein nspP123 containing nsP1, nsP2, and nsP3; polyprotein nspP1234 is produced by read-through of the opal codon (Feng et al., 1990). These polyproteins form the viral replication complex and are processed by the proteolytic activity of nsP2. In the late stage of infection, a positive-sense 26S subgenomic RNA is synthesized. Translation of the 26S RNA produces a structural polyprotein that is subsequently processed into individual structural proteins by a combination of viral and host proteases in the endoplasmic reticulum.

Structural characterization of alphaviruses includes cryoelectron microscopy image reconstructions of infectious particles determined to 9 Å resolution for Sindbis virus (Mukhopadhyay et al., 2006) and to 8.5 Å resolution for VEEV (Z. Li, personal communication). Crystal structures of the C-terminal domain of the capsid protein of Sindbis (Choi et al., 1991, 1996; Tong et al., 1992), Semliki Forest (Choi et al., 1997), and VEEV (S.J.W., unpublished data; Protein Data Bank [PDB] code: 1EP5) and the soluble ectodomain of Semliki Forest virus (SFV) E1 glycoprotein (Lescar et al., 2001) have been solved to atomic resolution. No structures have been solved of the alphavirus nonstructural proteins, but sufficient structural similarity with other proteins exists to permit construction of homology models for the central third of nsP2, the N-terminal domain of nsP3, and the central region of nsP4. The lack of high-resolution structures for alphavirus replication complex proteins prevents the use of powerful structure-based drug discovery and design methodologies to combat VEEV and alphavirus infections.

The alphavirus nsP2 protein has multiple enzymatic activities. A 456 amino acid N-terminal region (Gly1—Ile456, VEEV numbering) has been shown to possess ATPase and GTPase activity (Rikkonen et al., 1994), RNA helicase activity (Gomez de Cedron et al., 1999), and RNA 5′-triphosphatase activity (Vasiljeva et al., 2000). The 338 amino acid C-terminal region of nsP2 (Met457—Cys794, VEEV numbering) has been associated with regulating the 26S subgenomic RNA synthesis (Suopanki et al., 1998), downregulating minus-strand RNA synthesis late in infection (Sawicki et al., 2006; Sawicki and Sawicki, 1993), targeting nsP2 for nuclear transport (Peranen et al., 1999), and proteolytic processing of the alphavirus nonstructural polyprotein replication complex (Vasiljeva et al., 2001, 2003). Sequence analysis suggests that alphavirus nsP2 proteases are cysteine proteases and members of peptidase family C9 of clan CA (Rawlings et al., 2008). The nsP2 protease is an attractive target for antiviral therapeutics since it cleaves substrates with defined recognition sequences (Asp/Glu-Ala-Gly-Ala or Glu-Ala-Gly-Cys in VEEV) (Strauss and Strauss, 1994) and is required for alphavirus replication. Structural information for the nsP2 protease would greatly facilitate drug discovery and development efforts for VEEV and related alphaviruses. In
pursuit of this goal, we report the crystal structure of the C-terminal region of VEEV nsP2 (nsP2pro). This structure consists of two domains: a novel cysteine protease domain, followed by a methyltransferase-like domain of unknown function. Both residues of the catalytic dyad, Cys477 and His546, are located in the N-terminal domain, which is largely helical. The active site is positioned adjacent to the interface between domains, and both domains contribute to substrate recognition. However, the majority of residues involved in substrate binding are from the N-terminal domain. The effects of temperature-sensitive mutants in highly conserved residues identified in Sindbis and Semliki Forest virus alphaviruses (Agapov et al., 1998; Hahn et al., 1989; Lulla et al., 2006a; Suopanki et al., 1998) are explained. Also, explanations for the role of nsP2 in alphavirus RNA replication are proposed based on the structural similarity of the C-terminal domain to known RNA binding methyltransferases and the location in the structure of temperature-sensitive mutations known to affect RNA synthesis and processing.

Results and Discussion

Overall Structure

The VEEV nsP2 region chosen for structural studies was based on sequence alignment (Figure 1) between VEE and SFVs and on previous functional mapping of SFV
nsP2 (Vasiljeva et al., 2001) that defined a soluble nsP2 region with protease activity (nsP2pro; VEEV residues Met457–Cys794). This protein was expressed in E. coli, purified, and crystallized. Analysis of diffraction data indicated that nsP2pro crystals belonged to space group P2₁2₁2₁, and the crystal solvent content was determined to be 47.5% (34.0% acentric/centric data). Phasing power was 1.3808 Å, and the Matthews coefficient (Vm = 2.4 Å³/Da) (Matthews, 1968), each asymmetric unit was predicted to contain a single monomer of nsP2pro, and the protein correspond with regions of poor sequence conservation. The structure of the C-terminal domain of nsP2pro is 18%, with only 62 residues being conserved cysteine protease catalytic dyad formed by Cys477 and His546. It is organized around a central cluster of helices that are flanked by two short α helices. The catalytic histidine is part of a β strand. However, organization of the tertiary structure of the nsP2pro protease domain is different from that observed in papain and any other known protein structure, thus clearly indicating that the nsP2pro domain adopts a novel fold.

The C-terminal domain of nsP2pro extends from Arg604 to Ser793. This domain contains approximately equal amounts of helix and strand secondary structural elements arranged in three layers, with the faces of a central β sheet flanked by α helices. The function of the C-terminal domain is unclear, but 3D structural comparisons with DALLY (Holm and Sander, 1995), the NCBI Vector Alignment Search Tool (VAST) (Gibrat et al., 1996), and ProFunc (Laskowski et al., 2005) all indicate that the tertiary structure of the C-terminal domain of nsP2pro is similar to that of the methyltransferase family of enzymes (Figure 5). The structure of the nsP2pro C-terminal domain is similar to proteins belonging to the S-adenosyl-L-methionine (SAM)-dependent methyltransferases superfamly as defined by the SCOP (Structural Classification of Proteins) taxonomy (Murzin et al., 1995). This superfamly includes methyltransferases from a wide variety of organisms, including the flavivirus RNA cap (nucleoside-2'-O-)methyltransferase domain of RNA polymerase NS5 and the E. coli heat shock protein FtsJ RNA methyltransferase. Sequence identities are 21% between nsP2pro and FtsJ and 19% between nsP2pro and dengue virus NS5 methyltransferase. Sequence similarity between dengue virus NS5 methyltransferase and nsP2pro, which is believed to be enzymatically inactive, has been noted previously (Sawicki et al., 2006).

Table 1. X-Ray Crystallographic Data Collection and Processing Statistics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Native</th>
<th>KAu(CN)₃</th>
<th>Thimerosal</th>
<th>Terbium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>1.3808</td>
<td>1.5418</td>
<td>1.5418</td>
<td>1.5418</td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁2₁2₁</td>
<td>P2₁2₁2₁</td>
<td>P2₁2₁2₁</td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td>Unit cell parameters</td>
<td>a (Å)</td>
<td>47.5</td>
<td>47.3</td>
<td>47.5</td>
</tr>
<tr>
<td></td>
<td>b (Å)</td>
<td>72.6</td>
<td>72.7</td>
<td>72.7</td>
</tr>
<tr>
<td></td>
<td>c (Å)</td>
<td>106.7</td>
<td>105.6</td>
<td>105.6</td>
</tr>
<tr>
<td>Resolution limits (Å)</td>
<td>40.0–2.45</td>
<td>40.0–3.0</td>
<td>40.0–3.0</td>
<td>40.0–3.0</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.7 (100)</td>
<td>76.7 (81.0)</td>
<td>91.8 (95.7)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>R factor (working/free) (%)</td>
<td>19.3/24.7</td>
<td>0.15/0.19</td>
<td>0.15/0.19</td>
<td>0.15/0.19</td>
</tr>
<tr>
<td>Number of atoms (protein/water/formate)</td>
<td>2552/255/9</td>
<td>675/675/675</td>
<td>675/675/675</td>
<td>675/675/675</td>
</tr>
<tr>
<td>Average B factor (Å²)</td>
<td>31.4/34.5/59.1</td>
<td>31.4/34.5/59.1</td>
<td>31.4/34.5/59.1</td>
<td>31.4/34.5/59.1</td>
</tr>
<tr>
<td>Rmsd: bonds/angles</td>
<td>0.012/Å/1.43</td>
<td>0.012/Å/1.43</td>
<td>0.012/Å/1.43</td>
<td>0.012/Å/1.43</td>
</tr>
<tr>
<td>Ramachandran analysis (Non-Gly, Non-Pro)</td>
<td>Most favored 247</td>
<td>Allowed 33</td>
<td>Disallowed 1</td>
<td></td>
</tr>
</tbody>
</table>

Values for the highest-resolution shell are indicated by parentheses. All reflections were processed.

Table 2. Crystal Structure Refinement Statistics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution limits (Å)</td>
<td>40.0–2.45</td>
</tr>
<tr>
<td>R factor (working/free) (%)</td>
<td>19.3/24.7</td>
</tr>
<tr>
<td>Number of atoms (protein/water/formate)</td>
<td>2552/255/9</td>
</tr>
<tr>
<td>Average B factor (Å²)</td>
<td>37.1/34.6/59.1</td>
</tr>
<tr>
<td>Rmsd: bonds/angles</td>
<td>0.012/Å/1.43</td>
</tr>
<tr>
<td>Ramachandran analysis (Non-Gly, Non-Pro)</td>
<td>Most favored 247</td>
</tr>
</tbody>
</table>

Structure and Sequence Conservation

Sequence alignment across representative alphanavirus nsP2pro sequences reveals only limited sequence identity and moderate similarity amongst the related sequences (Figure 1). The catalytic residues Cys477 and His546 are invariant in all alphanavirus nsP2 sequences. The 3 residues immediately following each catalytic residue are also completely conserved, and at least 1, Trp547, has been shown to be necessary for proteolytic activity (Strauss et al., 1992). Overall sequence conservation in nsP2pro is 18%, with only 62 residues being invariant over all alphavirus strains. Interestingly, secondary structure elements (determined from the coordinates of the VEEV nsP2pro structure) in many regions of the protein correspond with regions of poor sequence conservation.

Proteolytic Domain and Catalytic Site

The core of the nsP2pro proteolytic domain consists of six helices flanked by two regions of short β hairpins.
and a single-turn 3_10 helix. Structure similarity searches with the NCBI VAST search (Gibrat et al., 1996), DALI (Holm and Sander, 1995), and the protein structure comparison service SSM (Krivine and Henrick, 2004) revealed very low similarity to several cysteine proteases, including papain, several cathepsins, and the FMDV leader peptidase. However, the results obtained with different structure comparison utilities were inconsistent. The majority of proteins that exhibited low similarity to the nsP2pro proteolytic domain were cysteine proteases, but their identity and statistical significance varied when analyzed by the different utilities. The DALI program identified the FMDV leader peptidase (PDB code: 1QMY) as having a low similarity to the nsP2pro domain, with a z score of 3.7 and 9% identity over 79 aligned residues. The VAST program identified papain (PDB code: 1PPN) and human cathepsin X (PDB code: 1EF7) as having little structural similarity to the nsP2pro domain. Cathepsin X was a slightly better structural match to the nsP2pro domain than papain since it was calculated to have a lower VAST p value and a higher residue percent identity within the superimposed regions. The regions of structural alignment between the VEEV proteolytic domain and cysteine proteases are very limited, and they include only a few residues in the immediate vicinity of the catalytic dyad. Other than nsP2pro, no known cysteine protease has been found to derive both residues of the catalytic dyad from the same domain. This suggests that the nsP2pro N-terminal domain is a novel cysteine protease fold.

The cysteine proteases identified as having low similarity to VEEV nsP2pro all belonged to the cysteine
proteinase superfamily (SCOP). These cysteine proteinases are characterized by a common catalytic core made of one α helix and three strands of β sheet (Murzin et al., 1995). In each of the cysteine protease structures, nsP2pro, papain, and human cathepsin X, the catalytic cysteine is situated at the N terminus of an α helix, and the catalytic histidine is located on a β strand (Figure 3). In papain and cathepsin X, the cysteine and histidine residues of the catalytic dyad are located in two separate domains, and the active site occupies the interface between these domains (Drenth et al., 1968; Smith, 1957). In contrast, in nsP2pro, the β strand containing the catalytic histidine is part of a short β hairpin within the N-terminal proteolytic domain and is not provided by a separate domain. Moreover, the nsP2pro secondary structure composition and topological arrangement differ significantly from cysteine protease structures deposited in the PDB. Notably, tertiary structure comparisons between nsP2pro and other proteins within the PDB indicate that, to our knowledge, the nsP2pro proteolytic domain represents a unique cysteine protease fold and a novel protein fold.

The nsP2pro structure confirms the hypothesis, generated from mutational studies (Strauss et al., 1992), that Cys477 and His546 come together to form a catalytic dyad within the protease active site (Figures 2A, 3A, and 4C). Although a conserved asparagine (Asn549) is located near these residues, and could potentially function as the third element of a catalytic triad, this asparagine residue is not oriented to interact with the histidine residue. Moreover, mutational studies of the related Sindbis virus (SINV) show that this residue is not essential for activity (Strauss et al., 1992).

A deep and pronounced groove transects the active site, suggesting the locations of S1, S2, and S3 sites in the protein that orient a peptide substrate relative to the catalytic dyad (Figure 4). A consensus peptide substrate Glu-Ala-Gly-Ala, corresponding to the nsP1-nsP2 cleavage motif, was modeled into the protease active site based on coordinates obtained by aligning Cys477, His546, and Trp547 with the corresponding residues in the Ulp1-SUMO complex (root-mean-square deviation [rmsd] of 1.1 Å) (PDB code: 1EUV). The backbone coordinates of the Ulp1 substrate bound to SUMO were used to position the P1–P4 residues of the nsP2 substrate consensus peptide. The P4 glutamic acid rotamer is also derived directly from this alignment. The position of the peptide substrate within the binding groove was based exclusively from the alignment of the active site residues of nsP2pro and Ulp1 and was not manually adjusted or refined by using molecular dynamic calculations. The fit of the Glu-Ala-Gly-Ala substrate within the nsP2 active site clearly delineates locations for the S1, S2, and S3 binding sites on the protease. These sites appear as shallow depressions on the protein surface, consistent with a surface that interacts with a peptide substrate containing either glycine or small side chain consensus residues. The S1, S2, and S3 sites line a long, deep groove formed at the interface between the nsP2Pro N- and C-terminal domains. Residues Asn544, Asn545, and His546 form a thumb that may regulate access into and out of this binding groove. The elevated temperature factors for residues within this region (Figure 2B) suggest that this protein segment is flexible. This thumb corresponds to a region of the protein where the SINV antigenic complex has 7 residues inserted in the aligned alphavirus sequences.

The S1 pocket is a small depression that is located ~6 Å from the catalytic cysteine and formed by residues Val476, Asn475, and Ala509. These S1 residues are either highly conserved or have only conservative substitutions (Figure 1). The backbone amides from Cys477 and Val476 likely form the oxyanion hole that is observed in many other cysteine and serine proteases.

The most significant contribution to defining the S2 binding site is made by Trp547; this site is conserved across all known alphavirus nsP2 sequences. The structures of several cysteine proteases that require substrates with a P2 glycine motif contain bulky aromatic residues immediately after the catalytic histidine, and these aromatic residues define the protease S2 site. Golubtsov and coworkers (2006) have called this the
glycine specificity motif (GSM). The nsP2pro structure shows that the conserved Trp547 indole nitrogen is positioned between two shallow surface depressions near the active site (Figure 4C). The interaction between the S2 tryptophan and the substrate P2 glycine suggests the use of the GSM selection strategy in alphavirus nsP2pros.

Residues Ile698 and Met702, located in the C-terminal domain, form the S3 binding site (Figure 3). Met702 is highly conserved (~75% identity) within the alphavirus strains listed in Figure 1. Although residue Ile698 is not highly conserved within the alphavirus strains, amino acid substitutions at this position are all hydrophobic, with methionine and isoleucine predominating. Additionally, the shallow S3 pocket is flanked by residues Ala509 and His510, both located at sites of high sequence conservation (~75%).

Three nsP2 cleavage sites are present in the nsP1234 polyprotein. All cleavage intermediates containing nsP2 have been shown to be proteolytically active (Vasiljeva et al., 2003). Protease activity in the polyprotein suggests the possibility of cis cleavage events contributing to processing as well as bimolecular trans cleavage. Available evidence is strongly suggestive that the processing of the nsP23 cleavage site occurs in trans (Vasiljeva et al., 2003). This suggestion is supported by examination of the nsP2pro structure. The measured distance from the last visible residue in the nsP2pro structure, Ser787, to the P1 alanine of the model substrate (representing the nsP23 cleavage site) is ~42 Å. A minimum of 12 residues in extended conformation would be necessary to span this distance. However, only 7 residues are missing at the C terminus of the nsP2pro structure, clearly indicating that the nsP23 cleavage site is not accessible to the protease active site in this conformation. This structural insight is in agreement with the proposed trans cleavage mechanism for nsP23 (Vasiljeva et al., 2003).

C-Terminal Domain

The fold of the SAM-dependent methyltransferase superfamily is described in SCOP (Murzin et al., 1995) as three layers, termed a/b/a, with a mixed β sheet of seven strands arranged in the order 3-2-1-4-5-7-6 and sandwiched between helices with strand 7 oriented antiparallel to the other strands. This is an appropriate description of the fold of the nsP2pro C-terminal domain, although one of the α helix layers consists of a single helix that is 5 residues long. This small helix results in many residues of β strands 6 and 7 being exposed to solvent in the nsP2pro structure. The nsP2pro C-terminal domain shows significant tertiary structure similarity to known methyltransferase structures (e.g., FtsJ, dengue virus NS5). An alignment between nsP2pro and FtsJ was
constructed by using only the six longest strands of the β sheet and the location of SAM in the FtsJ structure (Figure 5). All of the β strands align very well, and the long helix on the upper face of the β sheet is brought into close alignment. However, the backbone alignment in proximity to SAM is poor, and residues in the region of nsP2pro that correspond to the FtsJ methyltransferase SAM substrate binding site show no significant similarity to each other. Moreover, little sequence identity is observed in alphaviruses for residues aligned to the SAM substrate binding site. These observations are consistent with the proposal that the nsP2pro C-terminal domain lacks methyltransferase enzymatic activity despite having structural similarity to the FtsJ methyltransferase and low sequence similarity to dengue virus NS5 methyltransferase. However, the nsP2pro methyltransferase fold could be used as a scaffold to bind RNA elements that may regulate protease activity and virus replication.

**Alphavirus nsP2pro Functional Mutants**

Temperature-sensitive (ts) mutants in the C-terminal region of nsP2 that affect RNA synthesis and protease activity differently have been identified in related Sindbis and Semliki Forest viruses (Table 3) (Agapov et al., 1998; Hahn et al., 1989; Lulla et al., 2006b; Suopanki et al., 1998). We have examined the role of four temperature-sensitive mutations in the context of the VEEV nsP2pro structure (Figure 6). These temperature-sensitive mutations are of special interest because they occur at residues highly conserved across alphavirus strains and Semliki viruses and asparagines in other alphaviruses. Phe504 also interacts with Val477, which is not completely conserved, but is replaced by threonine in Aura virus. The location of Val477 near the protein surface suggests that threonine could be tolerated in this position by directing the hydroxyl group toward the surface and still presenting a nonpolar methyl group to Phe504. Mutation of Phe504 to leucine removes three nonpolar carbons from the interior of the N-terminal domain, leaving a void, which likely perturbs the hydrophobic core, resulting in instability at the nonpermissive temperature and loss of function.

The mutation of glycine to serine in ts24 could have multiple effects. The Gly723 main chain is completely buried and surrounded by the side chains of conserved residues Tyr784, Tyr724, His608, and Ala725. The backbone dihedral angles at this residue map to a region of the Ramachandran plot that is disallowed for nonglycine residues (lower-right quadrant of the phi-psi plot) (Ramachandran et al., 1983). Substitution of any residue here will strain the backbone and likely destabilize the protein. In addition, depending on the side chain rotamer, the surrounding residues would have steric clashes with a serine substitution at this position. These clashes and the energetic cost of burying a polar hydroxyl in the hydrophobic core would likely destabilize the hydrophobic interactions in the core and cause instability at the nonpermissive temperature and loss of function.

**Table 3. Alphavirus nsP2 Temperature-Sensitive Mutants**

<table>
<thead>
<tr>
<th>Mutant Name</th>
<th>Mutation (VEEV Residue)</th>
<th>Temperature-Sensitive Phenotype</th>
<th>Temperature-Sensitive Mutations</th>
<th>26S RNA Synthesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>ts7</td>
<td>D522N (N517)</td>
<td>D*</td>
<td>wt†</td>
<td>wt</td>
</tr>
<tr>
<td>ts18</td>
<td>F509L (F504)</td>
<td>wt</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>ts24</td>
<td>G736S (G723)</td>
<td>wt</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>S2</td>
<td>P726T (P713)</td>
<td>D</td>
<td>wt†</td>
<td>D</td>
</tr>
</tbody>
</table>

*“D” indicates a defect in normal function at the nonpermissive temperature. 
"wt" indicates normal wild-type behavior at the nonpermissive temperature.

**Figure 5. Superposition of the nsP2 C-Terminal Domain with E.coli FtsJ RNA Methyltransferase**

FtsJ methyltransferase is shown in green, and the C-terminal domain of VEEV nsP2pro is colored tan. SAM substrate bound to FtsJ is colored light blue. Superposition was obtained by maximizing the spatial overlap of the six longest β strands in the core region of each protein. Although the core β sheets are conserved, the remaining secondary structural elements are not. The secondary structure surrounding the SAM binding site is not conserved, indicating that this domain in nsP2pro is likely not a functional methyltransferase.
of nsP2 is organized into two discrete folded domains: an N-terminal domain that encompasses the protease catalytic dyad and active site, and a C-terminal domain with structural similarity to methyltransferases but with no known enzymatic activity. To our knowledge, the tertiary structure of the N-terminal domain has not been observed previously and represents a novel cysteine protease and protein fold. The substrate binding site and catalytic dyad are well defined. Placement of bound ligand from related cysteine proteases into the nsP2pro active site clearly delineates the S1, S2, and S3 binding sites and suggests substrate recognition and binding mechanisms. The C-terminal domain forms part of the active site and may be involved in substrate recognition, regulation of protease activity, and regulation of RNA replication. This structure will significantly aid drug discovery and development efforts to combat VEEV and related viruses.

Experimental Procedures

Cloning, Expression, Purification, and Crystallization of the Protease Domain of VEEV nsP2, nsP2pro

Cloning, expression, purification, and crystallization of nsP2pro have been described previously (Russo and Watowich, 2006). In brief, DNA coding for nsP2pro (residues Met457–Cys794 of VEEV nsP2) was amplified by polymerase chain reaction (PCR) and cloned into the pETBlue1 T7 expression vector (Novagen). Tuner DE3 (pLacI) E. coli cells were transformed with this vector, grown in flask cultures at 37°C, and induced by the addition of IPTG. Cells were pelleted and lysed, and the supernatant was processed through SP-Sepharose, Ni-Sepharose, and Superdex-200 chromatography columns to obtain >99% pure nsP2pro (data not shown). Mass spectrometry and N-terminal sequencing were used to verify the identity of the purified protein (data not shown). Purified nsP2pro was concentrated to ~6.0 mg/ml and crystallized with 3.0 M ammonium formate (pH unadjusted), 2% 2-methyl-2,4-pentanediol (MPD), 1% glycerol, and 0.2 mM zinc acetate.

Preparation of Isomorphous Heavy Atom Derivatives

Crystals were transferred from crystallization drops; washed in stabilizing solution containing 3.0 M lithium formate, 2% MPD, and 1% glycerol; and transferred to stabilizing solution containing 1–20 mM of the appropriate heavy atom compound. Crystals were soaked for a variety of times varying from 2 to 7 days in the heavy atom solution and then backsoaked into 2.5 M lithium formate, 2% MPD, and 40% glycerol. Typical crystals used for data collection were 400–500 μm long, 20 μm wide, and less than 5 μm thick.

Data Collection and Analysis

Crystals were soaked in a cryoprotectant solution containing 2.5 M lithium formate, 2% MPD, and 40% glycerol for 5–10 min before being flash cooled in a 100 K nitrogen gas stream. Diffraction data were collected at 100 K by using 1° wide frames on a DIP2030 imaging plate detector mounted on a MacScience M06HF rotating anode X-ray generator equipped with a 100 μm Cu Ka source and Rigaku confocal optics. A high-resolution native data set was collected on a MAR CCD at a wavelength of 1.3808 Å at the Center for Advanced Microstructures and Devices (CAMD) Gulf Coast Consortium Protein Crystallography PK1 beamline. Diffraction data were indexed, integrated, and scaled by using HKL2000 (Otwinowski and Minor, 1997). Identification of heavy atom sites, determination of phases, and density modification with DM (Cowtan, 1994) were performed by using the SHARP software package (Buster Development Group) (Bricogne et al., 2003) to give an initial experimental electron density map at 3.0 Å resolution. The initial Figure of Merit for acenatic (FOMac) and centric (FOMcen) reflections was 0.29 and 0.35, respectively. After density modification in DM, the mean FOM increased to 0.82. The DM-improved map was used for initial model building with TEXTAL (Romero et al., 2005); this placed polypeptide into >80% of the electron density map. Approximately half of the
residues placed were identified correctly. The model was improved with iterative rounds of manual model building in Xtalview (McRee, 1999), by using composite omit maps, and PDB/CNS (Brünger et al., 1998) refinement against the 2.46 Å data set. The stereochemical bond rmsd target was set to 0.012 Å, as determined by PDB (Singh et al., 2006) based on the ratio of observed:free parameters. Final refinement steps used the improved set of CNS bond length and angle parameters from the latest version of PDB. Sequence alignments based on secondary structure predictions were determined with SSAP (Krisinel and Henrick, 2004) accessed through the ProFunc web server (Laskowski et al., 2005).

Acknowledgments

We thank Dr. Henry Bellamy for assistance with X-ray diffraction data collection at the Center for Advanced Microstructures and Devices (CAMD) PX beamline. M.A.W. was supported by a grant from the Sealy and Smith Foundation to the Sealy Center for Structural Biology and Molecular Biophysics.

Received: June 27, 2006
Revised: July 29, 2006
Accepted: July 31, 2006
Published: September 12, 2006

References


Accession Numbers
Coordinates have been deposited in the Protein Data Bank (PDB) with accession code 2HWK.
Native Hepatitis B Virions and Capsids Visualized by Electron Cryomicroscopy

Kelly A. Dryden,1 Stefan F. Wieland,2 Christina Whitten-Bauer,2 John L. Gerin,3 Francis V. Chisari,2 and Mark Yeager1,4,*

1Department of Cell Biology
2Department of Molecular and Experimental Medicine
The Scripps Research Institute
10550 North Torrey Pines Road
La Jolla, California 92037
3Division of Molecular Virology and Immunology
Department of Microbiology and Immunology
Georgetown University Medical Center
Rockville, Maryland 20850
4Division of Cardiovascular Diseases
Scripps Clinic
10666 North Torrey Pines Road
La Jolla, California 92037

Summary

Hepatitis B virus (HBV) infects more than 350 million people, of which one million will die every year. The infectious virion is an enveloped capsid containing the viral polymerase and double-stranded DNA genome. The structure of the capsid assembled in vitro from expressed core protein has been studied intensively. However, little is known about the structure and assembly of native capsids present in infected cells, and even less is known about the structure of mature virions. We used electron cryomicroscopy (cryo-EM) and image analysis to examine HBV virions (Dane particles) isolated from patient serum and capsids positive and negative for HBV DNA isolated from the livers of transgenic mice. Both types of capsids assembled as icosahedral particles indistinguishable from previous image reconstructions of capsids. Likewise, the virions contained capsids with either T=3 or T=4 icosahedral symmetry. Projections extending from the lipid envelope were attributed to surface glycoproteins. Their packing was unexpectedly nonicosahedral but conformed to an ordered lattice. These structural features distinguish HBV from other enveloped viruses.

Introduction

HBV is a noncytopathic, enveloped, double-stranded (ds)DNA virus that causes acute and chronic hepatitis, and hepatocellular carcinoma (Chisari and Ferrari, 1995). Of the estimated 350 million people that are chronically infected with HBV, one million will die every year. In infected cells, transcription of the viral covalently closed circular (ccc)DNA template produces the pregenomic-(pg)RNA that is packaged with the polymerase into capsids formed by the core protein HBcAg (Figure S1 available in the Supplemental Data with this article online). Within the capsid, reverse transcription produces a single-strand (ss)DNA copy that serves as the template for second strand DNA synthesis. The resulting particles, containing dsDNA, are enveloped to become infectious virions (see Supplemental Data).

Bacterial expression of HBCAg results in spontaneous assembly of two types of capsids formed by 90 and 120 dimeric spikes assembled on icosahedral surface triangulation lattices of T=3 and T=4 symmetry, respectively (Böttcher et al., 1997; Conway et al., 1997; Crowther et al., 1994; Wynne et al., 1999). Milestones in the structural biology of HBV were high-resolution analyses by cryo-EM and X-ray crystallography of recombinant capsids, which showed that the capsid spikes are formed by 4-helix bundles comprised of dimers of α-helical coiled coils of HBCAg (Böttcher et al., 1997; Conway et al., 1997; Wynne et al., 1999). EM labeling studies have localized the amino terminus on the exterior of the capsid and the carboxyl terminus on the interior, which is also involved in nucleic acid binding (Conway et al., 1998; Zlotnick et al., 1997). Amino acids primarily at the base of the 4-helix bundle, as well as two acidic residues at the tip of the bundle, mediate envelopment (Böttcher et al., 1998; Ponsel and Bruss, 2003).

To date, very little is known about the structure and assembly of native capsids in infected cells (Kenney et al., 1995), and even less is known about the structure of mature virions. Here, we used cryo-EM and image analysis to examine native HBV capsids that do or do not contain viral nucleic acids and compare their structures to capsids within virions purified from patient sera.

Results

Capsid Purification and Morphology

Capsids from the livers of transgenic mice that replicate the virus (Guidotti et al., 1995) were purified by fractionation on cesium chloride density gradients. EM of negatively stained samples showed no distinction between the fractions other than concentration (Figure 1A). The majority of the particles were penetrated by stain, creating an “empty” appearance, with a subpopulation that excluded stain. Southern blot analysis with probes specific for HBV demonstrated that the majority of the viral DNA intermediates was associated with nucleocapsids of density 1.32 g/cm³, whereas capsids of lower density (1.29 g/cm³) did not contain any viral DNA (Figure 1B) or viral RNA (data not shown). These lower-density capsids were most likely released from nuclei that were disrupted during homogenization of the livers. This inference is based on the observations that capsids with the same low density can be isolated from the livers of transgenic mice that only express HBcAg and that the hepatocytes of these mice contain capsids only in the nuclei (Guidotti et al., 1994). The inability to purify HBV RNA containing capsids, even though they were present in the crude liver homogenate (data not shown), may indicate that immature, RNA-containing capsids are relatively unstable and/or transient particles.
Reconstructions of Cytoplasmic and Nuclear Capsids

To further examine the capsids that were positive and negative for HBV DNA, samples were analyzed by cryo-EM and image processing. Images of the two types of frozen-hydrated capsids were very similar in appearance (Figure 1C, top). The three-dimensional (3D) density maps of the native capsids at 14–16 Å resolution (Figure 1C, bottom) were also indistinguishable from the virus-like particles that assemble spontaneously upon in vitro expression of full-length or C-terminal, truncated HBcAg capsid protein (Böttcher et al., 1997; Conway et al., 1997; Wynne et al., 1999; Zlotnick et al., 1997). The T=4 capsids had a diameter of ~320 Å and were decorated with 120 spikes. In addition, internal density within the authentic capsids was similar to that seen for virus-like particles assembled by in vitro expression of full-length HBcAg, which encapsidate random cellular RNAs (Crowther et al., 1994; Zlotnick et al., 1997). Acquisition of random cellular RNA is also likely to occur for the capsids of 1.29 g/cm³ density, which show internal density that cannot be attributed to viral DNA (Figure 1C) or RNA (data not shown).

For the cytoplasmic, HBV DNA containing capsids (Figure 1C, bottom left), the internal density would correspond to one-third of the expected ssDNA genome, and we presume that the remainder is disordered. The internal density displayed in our maps at 14–16 Å resolution was even more apparent at 25 Å resolution (data not shown). However, capsid maps at higher resolution (<10 Å) did not show such density (Roseman et al., 2005). We presume that the DNA has some degree of low-resolution order, which is lacking at higher resolution. Disorder of encapsidated nucleic acid has also been observed with nodaviruses. At a resolution of 25 Å, maps of Flockhouse virus displayed a dodecahedral cage of duplex RNA (Tihova et al., 2004), whereas the crystal structure at 3.0 Å resolution only displayed ordered RNA at the icosahedral 2-fold axes (Fisher and Johnson, 1993). We infer that a portion of the HBV genome may assemble as DNA duplexes as a consequence of interactions with the arginine-rich inner surface of the HBcAg capsid (Birnbaum and Nassal, 1990). HBV DNA-containing capsids contain a single copy of the polymerase, which may play a role in forming a nucleation site for assembly in the infectious pathway.

T=3 and T=4 Symmetries of Cytoplasmic, Nuclear, and Virion Capsids

Previous studies have shown that about 10%–20% of the particles that assemble upon bacterial expression of HBcAg have T=3 rather than T=4 icosahedral symmetry (Crowther et al., 1994; Zlotnick et al., 1996). To date, it has not been known whether the T=3 particles are part of the normal viral life cycle. As shown in Figure 2A, we also observed T=3 sized particles derived from the livers of the HBV transgenic mice. These particles represented ~8% and 11% of the DNA-positive and DNA-negative capsids, respectively. Their 3D reconstruction displayed T=3 lattice symmetry and was nearly equivalent to that of the T=4 capsids.

Figure 1. Isolation of Native HBV Capsids and Analysis by Cryo-EM and 3D Image Processing

(A) Images of negatively stained fractions from the cesium chloride gradient.

(B) Southern blot of fractions probed for HBV DNA. Particles negative for HBV DNA (green outline) were in great excess and primarily represented capsids from the nucleus.

(C) There is close similarity between the 16 and 14 Å resolution 3D maps for capsids that are HBV DNA positive (blue outline) and negative (green outline), respectively (2-fold surface-shaded view [top], central cross-section view [middle], and surface view of the radii within the capsid [bottom]). The native particles encapsidate a dodecahedral cage of density ascribed to ~1 kb of ssDNA, which is closely associated with the inner wall of the capsid.

(D) Radial density plot of the two structures. The HBV-positive particles (blue) have slightly more density at low radii (<80 Å) than the HBV-negative particles (green). Bars, 500 Å.
to maps of particles assembled in vitro, with internal density close to the protein shell (data not shown).

Virions purified from the serum of HBV-infected patients also displayed two distinct size classes (Figure 2A). Radial density plots of the spherically averaged 3D reconstructions of each particle type clearly show that the smaller and larger enveloped virions have an inner profile that matches the T=3 and T=4 core particles, respectively (Figure 2B). The T=3 virions represented ~10% of the particles.

Equivalence of Capsid Architecture in Virions and Native Capsids

Three-dimensional image processing of the virion images was performed by separately examining the densities arising from the envelope and the capsid. The radial density profiles showed a close match for the peaks corresponding to the capsid shells within virions and the isolated cytoplasmic capsids (Figure 2B). This suggests that the capsid shell within virions maintains icosahedral symmetry, as demonstrated recently by analysis of T=4 capsids generated by stripping the envelope from virions (Roseman et al., 2005). Determination of orientation parameters for images of icosahedral viruses is enhanced by restricting the analysis to the annulus of densities relevant to the structure. To this end, we analyzed the capsids within T=4 virions by limiting the densities to radii spanning 90–170 Å (Figure 2B, top). The starting model was a reovirus core structure that was scaled to the capsid diameter with an artificial shell of density added to mimic the envelope (Figure 3A). Refinement of 178 images of the T=4 particles converged to a capsid structure that was similar to all T=4 capsid reconstructions to date (Figure 3A), including our maps of the cytoplasmic capsids (Figure 1C). This result implies that there were no gross structural rearrangements in the capsid during envelopment and virion secretion. Note that the map at the envelope radii was relatively featureless (Figure 3A).

A similar analysis of the T=3 virions was based on a limited number of particles (n = 21), due to their scarcity in the cryo-EM images. In this case, the starting model was based on a map of tomato bushy stunt virus (TBSV), scaled to the capsid diameter, with an artificial envelope shell. Image processing was restricted to radii spanning 90–155 Å (Figure 2B, bottom). When the 3D reconstruction was truncated at a diameter of 310 Å, the capsid displayed T=3 lattice symmetry (Figure 3B). As for the analysis of the T=4 virions, the envelope was relatively featureless. The unprocessed cryoimages (Figure 2A), the radial density profiles (Figure 2B), and the 3D reconstruction (Figure 3B) are the first evidence demonstrating the existence of T=3 HBV virions. Whether these T=3 enveloped particles are infectious remains to be determined.

Recently, cryo-EM and image analysis were used to analyze recombinant capsids formed by full-length HBCAg, as well as capsids derived by stripping mature virions with EDTA and dodecylmaltoside (Roseman et al., 2005). The recombinant capsids packaged random E. coli RNA, as well as some partially degraded HBCAg RNA. The overall structures of the RNA- and DNA-containing capsids were very similar. However, there were small shifts in density of a few angstroms throughout the capsid spike, in particular, additional density within a hydrophobic pocket at the base, which was thought to be important for envelopment. The resolution of our maps does not allow us to address this observation.

Image Analysis of Virion Envelope

The HBV envelope consists of hepatocyte membrane lipids and three viral surface glycoproteins (HBsAg) encoded by the same open reading frame, such that different start codons result in large (L), medium (M), and small (S) surface proteins (Ganem, 1996). It is noteworthy that the T=4 virions display some variation in diameter and deviation from a spherical shape. These features, as well as the heterogeneity in the L, M, and S proteins within the envelope, argue against strict icosahedral symmetry. Although there has been speculation on a one-to-one relationship between the core and surface glycoproteins, either as one L protein per spike as observed with peptide decoration of capsids (Böttcher et al., 1998) or as an S dimer per spike as proposed from the structural analysis of S HBsAg 22 nm particles (Gilbert et al., 2005), neither of these studies provides evidence that a 1:1 stoichiometry exists in vivo.

Nevertheless, to gain some insight into the general packing patterns of the proteins in the envelope, we first...
assumed that the particles have icosahedral symmetry. The T=4 particle data set was made as homogenous as possible by visually selecting the most spherical particles and including those with nearly identical diameters (n = 178) (Figure 2A). For image analysis of the virion envelope, eight different starting models were used, and the densities were restricted to radii spanning 170–250 Å. In Figure 3C, the starting model was the same as that used for analyzing the capsid radii (Figure 3A). The surface of the final map had distinct projections, but the capsid radii displayed a relatively smooth shell.

A similar analysis of the T=3 virions used TBSV as a starting model and was restricted to radii spanning 155–240 Å. The surface of the envelope in the final map displayed surface projections that were very similar in size and distribution to the T=4 virions (compare Figures 3C and 3D).

The 178 T=4 virion particles partially converged to several different yet similar structures. Representative results from three of the eight starting models are shown in Figure 4A. Map 1 is the same as shown in Figure 3C, map 2 arose from using the final map from analysis of the radii that spanned the capsid (Figure 3A) as the starting model, and map 3 arose from a starting model generated by applying random orientations to the images. For all reconstructions, the envelope density displayed surface projections of comparable size and spacing. However, there was clear variation between maps in the precise positions of the projections. For instance, in map 1, projections are centered on the 5-fold axes that are absent in maps 2 and 3, and in map 3, projections on the 3-fold axes are absent in maps 1 and 2 (Figure 4A). In addition, the observed lattices do not satisfy strict icosahedral symmetry, particularly evident as less distinct density on or around the 2-fold (map 1) and 5-fold (map 3) symmetry axes. Furthermore, specific triangulation symmetry could not be defined for any of the eight maps. The surface projections of the T=3 virion reconstruction (Figure 3D) were positioned most like those of map 2 but less distinct around the 5-fold axes. The density spanning the capsid radii displayed a smooth shell in all cases.

There were 160–200 surface projections in the various maps (Figures 3C and 4A), which we ascribe to the corrugated fringe of density, clearly visible on the perimeters of the raw images (Figure 2A, arrows). The projections extended ~30 Å from the envelope and were spaced ~60 Å apart (Figure 4B). On the basis of the observed volume of the projections and a protein partial specific volume of 0.74 cm³/g, an average projection corresponded to ~14 kDa of protein. The observed volume of a surface projection could account for the external domain of a single L or a glycosylated M protein. The packing patterns of the projections could arise if the 12 pentamers required to form a sphere (Ganser et al., 1999) are randomly inserted into a lattice, rather than at the precise 5-fold vertices necessary for an icosahedron.

The estimated volume of the envelope (~2.2 × 10⁷ Å³), excluding the density of the projections, could accommodate 370–400 S domains, which is the common motif for the three glycoproteins. This calculation assumes that the lipid partial specific volume is
0.98 cm³/g and that the virions contain 35%–40% lipid by weight, as reported for HBsAg 22 nm particles (Diminisky et al., 1997). The interpretation of our maps of HBV virions is that the surface projections are comprised of a random mixture of L and M domains, possibly paired with S proteins (Wunderlich and Bruss, 1996), which could contribute to their lack of icosahedral packing. The conformational switch of S dimers observed for 22 nm particles and proposed interactions with the underlying capsid have been suggested to bear on the formation of T=3 and T=4 virions (Gilbert et al., 2005). Our analysis of T=3 and T=4 virions shows no evidence of icosahedral order in the envelope or specific interactions with the capsid proteins. However, the low resolution of our maps does not allow us to exclude the possibility that the underlying S domains are icosahedrally ordered and only the L and M domains are more variably placed.

En Face Views Display Ordered Density
As a second, independent approach for examining the structure of the envelope, we analyzed en face views of T=4 virions, cytoplasmic capsids, and as a control, the capsid X-ray structure (PDB 1QGT) without any assumptions regarding the symmetry or packing of the surface projections. By extracting areas in the center of the particle, the radius of curvature is sufficiently small that at least at low resolution we can treat these areas as if they were 2D lattices. In addition, the particles were masked so as to exclude the circular edge of the capsid. En face images represent the superposition of the near and far sides of the virus, as well as any interior density. The three sets of images (virions, capsids, and the capsid X-ray structure) were classified independently by factor analysis. For all the particle types, the class averages displayed discrete densities (Figure 4C, left column), which gave rise to sampled reflections in the Fourier transforms (Figure 4C, middle column) and to peaks in circularly averaged 1D profiles of the Fourier transforms (Figure 4C, right column). Separate analyses of the capsid X-ray structure (Figure 4C, a) and the cytoplasmic capsids (Figure 4C, b) produced one predominant class, with a major peak at 1/43 Å in the 1D profiles arising from the distance between adjacent spikes. The virions yielded three primary classes (Figure 4C, c, d, e).

Figure 4. Image Analysis of HBV Envelope Surface Projections
(A) Maps with imposed icosahedral symmetry arising from the same virion data set but different starting models yielded variable yet comparably spaced surface features. Surface-shaded views of three representative maps (out of eight) viewed along the 2-fold (2F), 5-fold (5F), and 3-fold (3F) symmetry axes. Map 1, the same as shown in Figure 3C; map 2, result from the final map from refinement of density within radii spanning the capsid used as the starting model (Figure 3A); and map 3, result from a starting model generated by applying random orientations to the particles.
(B) Despite completely different starting models, each map displayed similar size and packing of the surface projections but had more poorly defined density at varying icosahedral symmetry positions. This pattern may arise if the projections have ordered packing with random insertion of pentamers. Packing of the surface projections is emphasized by the overlay of yellow dots. The red triangle has 60 Å edges and shows the average spacing of the projections.

Five-fold and 3-fold axes are indicated by black pentagons and triangles, respectively. (C) Analysis of en face images of HBV. Central, en face views of the T=4 capsid X-ray structure (a), cytoplasmic capsids (b), and virions (c, d, and e) were classified by factor analysis. For all the particle types, the class averages displayed discrete densities (Figure 4C, left column), which gave rise to sampled reflections in the Fourier transforms (Figure 4C, middle column) and to peaks in circularly averaged 1D profiles of the Fourier transforms (Figure 4C, right column). Separate analyses of the capsid X-ray structure (Figure 4C, a) and the cytoplasmic capsids (Figure 4C, b) produced one predominant class, with a major peak at 1/43 Å in the 1D profiles arising from the distance between adjacent spikes. The virions yielded three primary classes (Figure 4C, c, d, e).
and e), with peaks in the 1D profiles at \( \sim 1/45 \) and \( \sim 1/60 \) Å, which would arise from the capsid lattice (class I), the envelope projections (class II), and the superposition of both (class III). The fact that the envelope density cannot be refined to a unique icosahedral structure and that density can be interpreted as a 2D lattice from the en face analysis support our conclusion that the glycoprotein projections exhibit ordered, but nonicosahedral, packing with a separation of \( \sim 60 \) Å between surface projections.

**Discussion**

In this study, we used cryo-EM and image analysis to examine the native structure of HBV capsids from transgenic mice and virions isolated from patient sera. In the icosahedral analysis of the capsid within virions, the \( T=4 \) (Figure 3A) and \( T=3 \) (Figure 3B) capsids were well defined, but the envelopes were relatively featureless, and the projections visible in the raw images (Figure 2A, arrows) were not observed in the reconstruction. Likewise, in the analysis of the envelope, the internal \( T=4 \) capsid shells were relatively smooth, and the envelopes displayed distinctive surface projections (Figures 3C and 4). We attribute the blurring of the surface spikes within the capsid and the surface projections on the envelope to a lack of register between the internal capsid spikes and the virion envelope projections. That is, the capsid shell within the virion maintains the icosahedral structure observed in cytoplasmic (Figure 1C) or bacterially produced capsids (Böttcher et al., 1997; Conway et al., 1997; Crowther et al., 1994; Kenney et al., 1995; Roseman et al., 2005; Wynne et al., 1999), but the envelope and associated glycoproteins are not similarly ordered or in register with the underlying capsid. The packing pattern of the surface projections was also confirmed by analysis of en face views without any assumptions about symmetry (Figure S2). Our working 3D structure of the HBV virion is a composite model that combines the \( T=4 \) icosahedral capsid shell with the envelope decorated with 160–200 surface projections that are spaced \( \sim 60 \) Å apart (Figure 5A). Similar to the cytoplasmic capsids (Figure 1C), the analysis of the capsid shell within virions yielded an interior dodecahedral cage of density, which we ascribe to ordered mature dsDNA (Figure 5A, red) present in these virions as determined by Southern blot analysis (data not shown).

The radial density plots (Figure 2B) suggest that the capsid spikes are in close apposition but do not penetrate the inner leaflet of the lipid bilayer. This was confirmed by constructing a pseudoatomic model of the

![Figure 5. Model and Cartoon of HBV Virions](image)
capsid by docking the X-ray structure of recombinant capsids into the cryo-EM density map of the virion capsid (Figure 5B). If the interactions between the capsid spikes and the phospholipid headgroups were stabilized electrostatically, then one would expect a concentration of cationic amino acids at the tips of the spikes. Instead, there is a mixed distribution of charge (Figure 5B, right). This contrasts, for example, with the binding of nonmyristylated retroviral matrix proteins (MA) via electrostatic interactions between cationic residues in MA and anionic phospholipid head groups (Craven and Parent, 1996).

What then might direct virion assembly and stabilize the interactions between the envelope and the capsid? The L glycoprotein is required for capsid envelopment (Bruss, 1997) and has also been implicated in cell attachment and entry. About half of the L molecules undergo a surprising conformational change that translocates the pre-S, N-terminal domain to the external side of the envelope (Lambert and Prange, 2001). There is no density in our maps that can be ascribed to the ~50% of L domains on the capsid side of the bilayer, presumably because this portion of the protein is disordered or the protein is irregularly spaced. Nevertheless, mutational analysis of HBCAg identified 11 amino acids localized to a groove surrounding the base of the capsid spike (Figure 5B, green spheres) that allowed nucleocapsid formation but abrogated particle envelopment and virion formation (Ponsel and Bruss, 2003). Furthermore, peptides binding two amino acids (Glu77 and Asp78) near the spike tips (Figure 5B, red spheres) interfered with in vitro binding of the L-HBsAg (Böttcher et al., 1998). As shown in Figure 5B (right), the 160 residues of the loop of the L protein that interact with HBCAg are depicted as being unstructured because sequence analysis predicts almost no α helix or β sheet secondary structure. It appears then that the L molecules play important roles in capsid envelopment and in maintaining stable, noncovalent interactions between the capsid and envelope that generate virions.

In considering the structure of enveloped spherical viruses, we note that some, such as the alphaviruses, are strictly icosahedral with the same symmetry in the capsid and the envelope spikes (Zhang et al., 2002). In contrast, herpesvirus contains an icosahedral capsid (Zhou et al., 2002). In contrast, herpesvirus contains an icosahedral capsid (Zhou et al., 2002). HBV appears to be intermediate between these extremes. The capsid is icosahedral, and there is close association of the envelope and the surface of the capsid spikes. However, the molecular heterogeneity of the surface glycoproteins presumably interferes with their packing on an icosahedral lattice. The ordered but nonicosahedral packing of the surface glycoproteins is analogous to certain retroviruses such as murine leukemia virus that form capsids by random insertion of pentamers within a hexagonal lattice (Ganser et al., 1999).

**Experimental Procedures**

**Sample Preparation**

Details of the HBV-transgenic mouse lineage, capsid isolation, DNA analysis, and virion preparation are described in the Supplemental Data. Briefly, for capsid isolation, the livers of transgenic mice (Tg[HBV 1.3 genome]Chi46), which were positive for HBeAg, were homogenized and enriched by centrifugation. An aliquot was re-moved for Southern blot analysis, and the remainder of the supernatant was layered onto a sucrose cushion. The pellets were suspended in CsCl buffer, and after ultracentrifugation, the fractions from four gradients were collected and pooled and further concentrated in BioMax filters. The filtrates were loaded on a second CsCl gradient, the density of the collected fractions was measured via the refractive index, and the fractions were diluted in PBS and pelleted by centrifugation. Of the resulting 30 µl, 10 µl was removed for Southern blot analysis and the remainder used for EM.

Virions were obtained from multiple plasmapheresis units from a single HBV-positive individual that were purified by multiple steps of centrifugation (Kaplan et al., 1973). Samples enriched in virions were resuspended in PBS to a final concentration of 50 x compared with the original starting material, divided into 50 µl aliquots, and stored at ~80°C.

**EM**

All capsid fractions and virion samples were negatively stained with 2% uranyl acetate on continuous carbon grids and examined by EM using a Philips CM100 (Philips/FEI, Eindhoven) operating at 100 kV. Selected samples were vitrified by standard methods for cryo-EM (Yentsch et al., 1994). In brief, a sample was placed on a carbon-coated grid, blotted, and rapidly plunged into liquid ethane.

**Image Processing**

Micrographs with minimal astigmatism and drift were digitized with a Zeiss microdensitometer (ZI Imaging), corresponding to 3.0 Å at the level of the specimen. Particle images were extracted with X3D (Conway et al., 1996) and were processed by polar Fourier transform methods (Baker and Cheng, 1996). The contrast transfer function was corrected by using parameters calculated from the program Bsoft (Heymann, 2001). For the analysis of the T=4 and T=3 cytoplasmic capsids, the starting model was a density map of the reovirus core (Dryden et al., 1993) or TBSV (http://mmtsb.scripps.edu/viper), respectively, each rescaled to match the diameter of the HBV capsid. For virions, the same starting models were used but with an artificial shell of density added to mimic the envelope (Figure 3). The ability to determine correct orientation parameters for each particle image is enhanced when the analysis is restricted to an annular window of the density, centered on the particle. This was critical for analysis of the HBV virions. For orientation analysis of the capsids within virions, the densities were limited to radii spanning 90–170 Å and 90–155 Å for the T=4 and T=3 virions, respectively. For analysis of the virion envelope, seven additional starting models were used for the T=4 virions, and the densities were restricted to radii spanning 170–250 Å (Figures 3C and 4). Only the TBSV starting model was used for the T=3 virions (Figure 3D), and the densities were restricted to radii spanning 155–240 Å. For the capsid and virion reconstructions, all correlation coefficients were greater than 0.45 and 0.35, respectively. Fourier shell correlations of the reconstructions using a cut-off value of 0.5 yielded estimated resolutions of 16, 14, and 28 Å for the DNA-positive capsids, DNA-negative capsids, and virions, respectively.

Analysis of 2D en face images was performed with EMAN software (Ludtke et al., 1999). An area of 180 × 180 Å at the center of each image was masked so as to exclude the circular edge of the capsid. Because the virions were the limiting data set, an equivalent number of images (178) were selected of the cytoplasmic capsids. As a control experiment for the en face analysis, we used the program pdb2mrc to generate a low-pass filtered map of the capsid X-ray structure (PDB 1QGT) at 20 Å resolution to which 2-sigma noise was added. We then generated 178 random projections of this low-resolution X-ray map from which we masked central areas of 180 × 180 Å. The three sets of particle images (virions, cytoplasmic capsids, and the capsid X-ray structure) were separately classified into groups by factor analysis using Starrclasses. The images were iteratively aligned, averaged, and assigned to classes by crosscorrelation using Classsebymra. Fourier transforms and their 1D profiles were calculated with ctffit.

An icosahedral construct of the recombinant HBCAg capsid was calculated from the X-ray crystal structure by using Viper and visually docked into the cryo-EM density map of the virion capsid.


Supplemental Data

Supplemental Data include Supplemental Experimental Procedures, Supplemental References, and one figure and can be found with this article online at http://www.molecule.org/cgi/content/full/22/6/843/DC1/.

Acknowledgments

We thank Barbie Ganser-Pornillos for helpful discussions and Michael E. Pique for assistance with preparation of figures. This work was supported by National Institutes of Health grants RO1 GM066087 (to M.Y.), and RO1-CA40489 and RO1-AI20001 (to F.V.C.).

Received: January 17, 2006
Revised: March 18, 2006
Accepted: April 25, 2006
Published: June 22, 2006

References


