Binding of a neutralizing antibody to dengue virus alters the arrangement of surface glycoproteins


The monoclonal antibody 1A1D-2 has been shown to strongly neutralize dengue virus serotypes 1, 2 and 3, primarily by inhibiting attachment to host cells. A crystal structure of its antigen binding fragment (Fab) complexed with domain III of the viral envelope glycoprotein, E, showed that the epitope would be partially occluded in the known structure of the mature dengue virus. Nevertheless, antibody could bind to the virus at 37 °C, suggesting that the virus is in dynamic motion making hidden epitopes briefly available. A cryo-electron microscope image reconstruction of the virus:Fab complex showed large changes in the organization of the E protein that exposed the epitopes on two of the three E molecules in each of the 60 icosahedral asymmetric units of the virus. The changes in the structure of the viral surface are presumably responsible for inhibiting attachment to cells.

Dengue, yellow fever and West Nile viruses are major human pathogens that are members of the flavivirus genus of the Flaviviridae family. Dengue virus (DENV) is the causative agent for dengue fever and the more severe dengue hemorrhagic fever (DHF). Although DENV infects approximately 50–100 million people each year, no effective vaccine has been licensed for human use. Vaccine development has been hampered by the potential complications following secondary DENV infections, which can result in DHF. This frequently occurs when the secondary infection is of a serotype different from the first, possibly because cross-reactive non-neutralizing antibodies from the first infection bind to the virus and promote antibody-dependent enhancement (ADE) of infection in cells expressing Fc-γ receptors. To avoid ADE, a tetravalent vaccine would need to elicit strongly neutralizing antibodies against all four serotypes of DENV.

Flaviviruses consist of an icosahedrally symmetric ectodomain, containing 180 copies of the envelope (E) glycoprotein and 180 copies of the membrane (M) protein anchored in and surrounding a lipid membrane. The nucleocapsid core within the membrane consists of a positive-sense, 11-kb RNA genome and multiple copies of the capsid protein. In the mature virus, the E protein is arranged into 30 rafts of three parallel dimers. Crystal structures of the dimeric E protein have been determined for DENV serotypes 2 and 3. The E-protein monomer has three domains, E-DI, E-DII and E-DIII, of which E-DII is probably involved in recognition of the principal cell receptor. However, E-DII can also participate in the initial binding to an ancillary receptor, DC-SIGN, thereby possibly enhancing the local concentration of the primary receptor in the vicinity of the virus required for cell entry.

Neutralizing epitopes are clustered at the tip of E-DII (which is also the location of the fusion peptide), the hinge region between E-DI and E-DII, and the lateral surface of E-DIII. The hinge regions, between E-DI and E-DII and between E-DI and E-DIII, participate in structural rearrangements that occur at low pH as the immature virus is converted into infectious particles and in the initial stages of infection when the virus fuses with an endosomal membrane.

Before the fusion of the virus to endosomal membranes, the E dimers dissociate and then reassociate as trimers. This conformational change can be blocked in West Nile virus (WNV) infection by the monoclonal antibody (mAb) E16 binding to E-DIII. The postfusion, trimeric E-protein structures of tick-borne encephalitis virus and DENV show that E-DIII rotates by about 70° closer to E-DI relative to its position in the dimeric prefusion structure.

The mAb 1A1D-2 strongly neutralizes DENV serotypes 1, 2 and 3 (Supplementary Fig. 1 online) but does not bind to serotype 4 (ref. 21). We report here the crystal structure of the Fab fragment of the mAb 1A1D-2 complexed with recombinant E-DIII of DENV serotype 2. Interpretation of a cryo-electron microscopy (cryoEM) image reconstruction of Fab 1A1D-2 complexed with DENV, made by using this crystal structure, showed that Fab 1A1D-2 bound to only

1Department of Biological Sciences, Purdue University, 915 West State Street, West Lafayette, Indiana 47907-2054, USA. 2Department of Pathology & Immunology and 3Department of Medicine, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, Missouri 63110, USA. 4Arbovirus Disease Branch, Division of Vector-Borne Infectious Diseases, Centers for Disease Control and Prevention, Public Health Service, US Department of Health and Human Services, P.O. Box 2087, Fort Collins, Colorado 80522, USA. 5Department of Molecular Microbiology, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, Missouri 63110, USA. 6Present address: Solid State Chemistry, An Aptuit Company, 3065 Kent Avenue, West Lafayette, Indiana 47906, USA. Correspondence should be addressed to M.G.R. (mrg@purdue.edu).

Received 24 September 2007; accepted 4 January 2008; published online 10 February 2008; doi:10.1038/nsmb.1382
two of the three E proteins in each of the 60 icosahedral asymmetric units and resulted in the rearrangement of E proteins, thus preventing viral infection.

RESULTS

Crystal structure of the Fab–1A1D-2: E-DIII complex

The recombinant E-DIII protein of DENV serotype 2 was complexed with Fab 1A1D-2. The structures were determined for two crystal forms of the complex (Table 1 and Supplementary Fig. 2a online). Both forms had one complex per crystallographic asymmetric unit. Crystal form 1 diffracted to only 3.8-Å resolution, whereas form 2 diffracted to 3.0-Å resolution. The r.m.s. deviation between equivalent Cz atoms of the two structures using only the constant domain of Fab, the variable domain of Fab or E-DIII were 1.2 Å, 1.7 Å and 1.2 Å, respectively. In addition, the E-DIII structure in the complex was similar to the E-DIII structure in crystallized DENV E protein6,8, with an r.m.s. difference of only 1.1 Å between equivalent Cz atoms.

The main difference between the two crystal forms is a 20° change in the elbow angle of the Fab molecule (Supplementary Fig. 2b), consistent with previous observations that the hinge angle in Fab fragments is flexible19,20. The surface area of the interface between the Fab molecule and E-DIII is 940 Å², typical of antibody-antigen interactions25. The binding surface on the Fab molecule consists of five of the six complementary determining regions L1, L2, H1, H2 and H3. The E-DIII binding surface is predominantly on one β-strand (residues 305–312), but residues 325, 364, 388 and 390 also contact the Fab molecule (Fig. 1a and Supplementary Table 1a online). There are likely to be about eight hydrogen bonds, three salt bridges and some hydrophobic interactions in the interface. The relative importance of the residues in the contact area was examined by investigating whether site-specific mutant forms of E-DIII (Supplementary Fig. 3 online) could bind to mAb 1A1D-2 using a yeast surface-display system. Mutations of K305E, K307E or K310E completely abolished antibody

Figure 1 Interactions of E-DIII with the Fab 1A1D-2 molecule and comparison of Fab 1A1D-2 binding to DENV E-DIII with Fab E16 binding to WNV (Protein Data Bank accession code 1ZTW)26. (a) Stereo diagram of the interface between E-DIII and Fab 1A1D-2 in the crystal structure complex. Residues of Fab 1A1D-2 and E-DIII are colored yellow and cyan, respectively. The light (L) and heavy (H) chain of the Fab molecule are indicated in the residue label with subscripts. Oxygen and nitrogen atoms are colored red and blue, respectively. Putative hydrogen bonds are shown as dotted lines. (b) Footprints of Fab 1A1D-2 (cyan) and E16 (pink) on E-DIII. Residues recognized by both Fab are colored purple. (c) Ribbon diagram of the E molecule onto which the E-DIII components of the Fab 1A1D-2 and E16 complex crystal structures have been superimposed. E-DI, E-DII and E-DIII are colored red, yellow and blue, respectively. The Fab 1A1D-2 (cyan) and Fab E16 (pink) molecules are shown as Cα backbones. (d) The pseudo atomic–resolution structure of the E-protein arrangement in mature dengue virus (DENV)31. The E-protein molecules at the A, B and C sites in the asymmetric unit of the virus are shaded.

Figure 2 Fab 1A1D-2:DENV complex formation is temperature dependent. (a) DENV with no antibody. (b) DENV incubated with Fab 1A1D-2 at room temperature showed few particles with bound Fab molecules. (c) DENV incubated with Fab 1A1D-2 at 37 °C showed that most particles were saturated with Fab molecules.
binding, consistent with the crystal structure, which showed that these residues participate in making salt bridges and hydrogen bonds with the Fab molecule.

The importance of the β-strand, formed by residues 305–312 (β-strand A3), to the binding of various other neutralizing antibodies against DENV is evident in that several escape mutations to neutralizing antibodies are located within the strand. This region is also recognized by the antibody 4E11, which cross-reacts with and neutralizes all DENV serotypes.

Comparison of the amino acids in the 1AID-2 DENV2 epitope with other serotypes (Supplementary Table 1b calculated using the program ClustalW, http://www.ebi.ac.uk/Tools/clustalw/), indicated that the DENV2 epitope is closely similar to DENV1 and DENV3, but not DENV4 (similarity score of 45, 45 and 18, respectively). This is consistent with the inability of the antibody to bind DENV4 (ref. 21).

Comparison of mAb 1AID-2 and E16 binding sites on E-DIII

The neutralizing mAb E16 inhibits WNV infection at a step after attachment, but before fusion, whereas the mAb 1AID-2 neutralizes DENV infection of Vero cells, in part by preventing viral attachment. Although both antibodies bind E-DIII, their footprints barely overlap one another (Fig. 1b,c). Presumably, this difference, at least in part, is likely to account for the difference in their mechanisms of neutralization. The surface accessibility of the E16 and 1AID-2 epitopes on E-DIII, as determined by crystallography, are different for each of the three E molecules in the native virus icosahedral asymmetric unit, identified as A, B and C in Figure 1d. The epitope recognized by E16 has an area of 1,550 Å² and is fully exposed on E-DIIIIB (E-DIII at B site) and E-DIIIC (E-DIII at C site) of the mature virus, whereas only 54% of the epitope is exposed for the potential binding site on E-DIIIA (E-DIII at A site). This is consistent with the cryoEM reconstruction map of Fab E16 complexed with WNVL, which shows Fab molecules binding only to E-DIIIIB and E-DIIIC, but not to E-DIIIA.

In contrast to Fab E16, 18% of the binding surface of Fab 1AID-2 is buried at all three E-DIIIs in the icosahedral asymmetric unit of the mature virus. This raises the question of how the antibody could bind to any of the E-DIII epitopes. Temperature was found to be crucial in the binding of the antibody (Fig. 2). Inspection of cryoEM micrographs showed that fewer than one-third of the particles were bound by Fab 1AID-2 when incubated for 30 min at room temperature, whereas nearly all the particles bound Fab at 37 °C. This suggested that higher temperature promotes increased mobility of the E proteins on the surface of the virus, thus transiently exposing the previously hidden part of the epitope and making it available for antibody binding.

Structure of dengue virus 2 complexed with Fab 1AID-2

The cryoEM reconstruction of Fab 1AID-2 complexed with DENV2 incubated at 37 °C was determined to 24 Å resolution (Fig. 3). Interpretation of the cryoEM map showed that the E glycoproteins on the viral surface had undergone a major rearrangement (Fig. 4). Such large quaternary structural changes are not uncommon in flaviviruses, as occurs in the maturation of virions and in the fusion of virions to cell membranes.

The interpretation of the cryoEM map (Fig. 3a) was initially guided by the two large, flat features in each icosahedral asymmetric unit, presumably representing the Fab molecules. Therefore, the first fitting operation was to manually position the 1AID-2-E-DIII crystal structure into the appropriate densities. This showed that the elbow angle found in crystal 2 (Supplementary Fig. 2b) correlated better with the cryoEM density at both of these sites. The initial manual fitting results were optimized using the EMfit program (Supplementary Fig. 4 and Supplementary Table 2b online). The positions of all E-DIII domains in the antibody complex were similar to those in the native virus (Figs. 1d and 3b). Location of E-DI-DII at sites A and C (Fig. 3b) in the cryoEM map were predicted by superimposing E-DIII of the E protein as found in the mature virus onto the already positioned E-DIIs complex structure. The E-DI-DII at sites A and C then had to be translated about 15 Å and 8 Å, respectively, to satisfy the nearest densities. The remaining uninterpreted density corresponded to the E molecule at site B (Supplementary Table 2b).

The resultant structure had the parallel E molecules at sites A, B and C in the mature virus (Fig. 1d) rotated counterclockwise (viewed from outside the virus) by roughly 33°, 32° and 73°, respectively (Fig. 3b). As a result, the epitopes on E-DIIIA and E-DIIIB had become completely exposed and their associated E-DIIs had their fusion tip pointing toward the viral lipid envelope (Fig. 3b and Supplementary Fig. 5a-c online). The E molecules at site C seem to make dimers by pairing across the icosahedral two-fold axes (Supplementary Fig. 5b). The E protein at site B had its Fab binding epitope turned inwards. Comparison of the E-protein structures at the A, B and C sites with the structure of the E protein in the mature virus showed that the unliganded E molecule corresponding to the B site had changed little compared to its structure in the mature virus. However, the structure of the other two E molecules with bound Fab molecules at sites A and C caused a change in the position of E-DIII relative to E-DI-DII (Supplementary Fig. 5d).
DISCUSSION

The ability to expose the partially hidden epitope at all E-DIII sites can be explained if it is assumed that icosahedral viruses can engage in ‘breathing’ motions. Dynamic structural changes have been demonstrated for icosahedral nodavirus\(^{22}\) and rhinovirus\(^{23}\). The breathing exposes internal regions of rhinovirus structural proteins (VP1, VP2 and VP3), making them available for protease cleavage. The motion can be inhibited by certain antiviral compounds that bind into a cavity within VP1 (ref. 33), thereby stabilizing the virus and reducing the extent of the breathing. Apparently, in the present case, the breathing of DENV can expose the 18% of the hidden 1AID-2 epitope for Fab attachment when incubated at an elevated temperature (Fig. 2), but is insufficient to expose 46% of the hidden surface at the WNV E-DIIIA site to allow binding of Fab E16 (ref. 19). The ability of Fab 1AID-2 to bind to epitopes near the five-fold (the E-DIIIA site) and three-fold (the E-DIIIC site) vertices, but not elsewhere on the viral surface, suggests that the extent of the breathing motion might be greatest at these vertices.

Although, presumably, breathing exposes the epitopes required for 1AID-2 binding to DENV, breathing also implies that, in the absence of antibody, the E proteins return to their original positions in the mature virus. Once a Fab molecule has bound, the virus would probably not be able to return to its previous range of dynamic motions because of steric hindrance between the bound Fab molecules and neighboring E proteins. Thus, the Fab molecules may capture transient viral intermediate conformations. However, there is a problem with this concept. If the observed structure were in the normal dynamic range of the viral conformations, then E16 should have captured the same state, as its epitope on E-DIII near the five-fold vertices is completely exposed in the cryoEM structure of the 1AID-2: virus complex. However, E16 Fabs did not bind to these sites, even at 37 °C (B. Kaufmann, Purdue University, unpublished data). Thus, possibly, once a 1AID-2 Fab molecule has bound to a transient intermediate conformation, the E proteins plus bound Fab molecules move to positions that do not occur during the normal breathing of the native virus to relieve the steric stress. One caveat is that this analysis assumes that DENV and WNV have the same breathing motion, because E16 binds to WNV and not to DENV. Another possibility is that, once E16 has bound to the fully available B and C sites, then breathing might be diminished or altered, making binding of E16 to any other normally hidden epitope impossible.

Thus, the structure of the 1AID-2:virus complex might be a trapped form of a breathing mode rather than a dead-end conformation.

If it is assumed that the particle will tend to conserve its icosahedral symmetry during breathing, then as soon as one antibody has bound, the rest of the virus alters its structure to the antibody-bound conformation. Thus, the binding of one or only a few 1AID-2 mAbs will make all other E proteins more accessible for antibody binding. Hence, a low concentration of antibody would be sufficient to alter the surface structure of the virus. This cascade mechanism is supported by the observation that some particles seen on micrographs of DENV incubated with Fab 1AID-2 at room temperature were covered by Fab molecules, whereas others had none (Fig. 2).

The mAb 1AID-2 probably inhibits DENV attachment by several different mechanisms. One block to infection could be the result of the antibody altering the spatial distances between the glycans on the E proteins and, therefore, inhibiting the interaction of the virus with its ancillary attachment receptor, DC-SIGN\(^{13,34}\). Another block to infection may be that the antibody binding to DIII of the E glycoproteins prevents binding of the virus to its primary entry receptor\(^{21,12}\).

An understanding of the mechanism of neutralization of an antibody could elucidate its likelihood for promoting ADE at a given concentration. All neutralizing antibodies can promote ADE \textit{in vitro} at subneutralizing concentrations\(^{15-18}\). Recent stoichiometric analysis of flavivirus antibody binding to virus suggests that there is a threshold occupancy required for neutralization\(^{20}\). Concentrations of antibody that fall below this threshold or antibodies that never reach this threshold, even at maximal binding, will enhance infection \textit{in vitro} in cells expressing Fc-γ receptors. The mechanism of antibody neutralization may determine the occupancy requirement for inhibition of infection and, therefore, influence the range of concentration that promotes ADE. Antibodies that neutralize solely by sterically blocking receptor attachment may require higher concentrations for complete neutralization and, therefore, have a greater range of concentration for sustaining ADE. Indeed, this was recently observed with DII-specific mAbs against WNV\(^{14}\). In contrast, mAbs such as 1AID-2 that promote a cascade of E-protein rearrangements on DENV may have lower occupancy requirements for neutralization. Vaccines that consist of an epitope recognized by mAb 1AID-2 may thus elicit an antibody response with greater inhibitory activity and less possibility of enhancing infection over a wider range of antibody concentration.

METHODS

Neutralization assay. To determine the concentration of mAb required to reduce the number of plaques on BHK-21 cells, increasing concentrations of mAb 1AID-2 were added to 50 plaque-forming units (p.f.u.) of DENV and incubated at 37 °C for 1 h. We then added 300 μl of the mixture to a monolayer of BHK-21 cells in a 6-well plate and incubated for 1 h at 37 °C. Supernatant was removed and 3 ml of 1.0% (w/v) carboxyl methyl cellulose in MEM plus 5% (v/v) FCS was layered onto the infected cells. After further incubation at 34 °C for 8 d, the wells were stained with 0.5% (w/v) crystal violet dissolved in 25% (v/v) formaldehyde to visualize the plaques. Percentage of plaque reduction was calculated as:

\[
\text{Percent plaque reduction} = 100 - \left( \frac{\text{plaque number when incubated with mAb}}{\text{plaque number without mAb}} \right) \times 100
\]

1AID-2 neutralized DENV serotypes 1, 2 and 3, although DENV3 was least sensitive to this antibody (Supplementary Fig. 1).

Crystallization of the 1AID-2 Fab:E-DIII complex. A two-fold molar excess of E-DIII (Supplementary Methods online) was added to 1AID-2 Fab (Supplementary Methods) and incubated at room temperature for 2 h. The complex was purified with a Superdex 75 column (GE Healthcare) in 20 mM HEPES, pH 7.5, 150 mM NaCl and 0.01% (w/v) sodium azide, and concentrated to 2 mg ml\(^{-1}\). Successful crystallization conditions were 0.1 M MES...
Table 1 Data collection, processing and refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>Crystal form 1</th>
<th>Crystal form 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
<td>C2</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>164.78, 66.89, 79.41</td>
<td>168.12, 58.65, 78.20</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90, 113.1, 90</td>
<td>90, 114.4, 90</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>3.8</td>
<td>3.0</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.118 (0.198)</td>
<td>0.107 (0.28)</td>
</tr>
<tr>
<td>I / σI</td>
<td>6 (3)</td>
<td>11 (3.5)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>74.25 (54.3)</td>
<td>98.5 (90.1)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>2.3 (1.7)</td>
<td>3.4 (2.8)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>20–3.8</td>
<td>20–3.0</td>
</tr>
<tr>
<td>No. reflections</td>
<td>58,817 (509)</td>
<td>13,829 (1,553)</td>
</tr>
<tr>
<td>Rwork / Rfree</td>
<td>31/36</td>
<td>26/32</td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>3,888</td>
<td>3,980</td>
</tr>
<tr>
<td>Water</td>
<td>0</td>
<td>254</td>
</tr>
<tr>
<td>β-factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>20.0º</td>
<td>39</td>
</tr>
<tr>
<td>Water</td>
<td>20.0º</td>
<td>34</td>
</tr>
<tr>
<td>r.m.s. deviations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>2.69</td>
<td>2.27</td>
</tr>
</tbody>
</table>

Values in parentheses are for highest-resolution shell.

*p*-Factor value is set to 20 and not refined because of low-resolution data.

pH 5.8, 12% (w/v) PEG 3350 and 0.1 M MES, pH 5.4 and 11–13% (w/v) PEG 2000. Both kinds of crystals were needle-shaped with maximum dimensions of 100 × 30 × 30 µm. The crystals were frozen using 25–28% (w/v) PEG 400 as a cryoprotectant in addition to the mother liquor.

Crystallographic structure determination. Diffraction data of both crystal forms were collected using a wavelength of 0.979 Å at 100 K on beamline 23ID at the Advanced Photon Source. The data were indexed and scaled using HKL2000 (ref. 39; Table 1). Both crystal forms belong to space group C2, differing primarily in only a 10% change of the monoclinic b axis. The structures of both crystal forms were solved by molecular replacement using the program MOLREP40. Several immunoglobulin structures were used as search models before a good solution was attained for each of the crystal forms (PDB code 1A3R and 1BBD for crystal forms 1 and 2, respectively). The trial structures were initially refined as rigid bodies using the CNS program system41. Electron density corresponding to E-DIII was found to be adjacent to the hypervariable region, as expected. The E-DIII and Fab 1AID-2 molecules (Supplementary Methods) were modeled into the electron density using the program O42. Coordinates were refined with the program CNS43 using reflections between 20-Å and 3.8-Å spacing for crystal form 1 and between 20-Å and 3.0-Å spacing for crystal form 2 (Table 1).

Mapping of mAb 1AID-2 epitope by yeast surface display. We expressed DENV2 (strain 16681) E-DIII (residues 296–415) of E in yeast43 by engineering BamHI and Xhol restriction enzyme sites at the 5′ and 3′ ends of the E-DIII gene, using PCR amplification from an infectious cDNA clone43. This fragment was cloned into the BamHI and Xhol sites of the pYD1 vector (Invitrogen) and expressed in the Saccharomyces cerevisiae strain EBY100. Single point mutations at residues Thr303, Lys305, Lys307, Lys310 and Asn390 were made using the QuikChange mutagenesis kit (Stratagene). Yeast that expressed mutant serotype 2 DENV E proteins was incubated with 50 µl of mAb (25 µg ml−1) on ice for 30 min. The yeast was washed three times with PBS supplemented with 1 mg ml−1 BSA, incubated with a 1:500 dilution of a goat anti-mouse IgG that had been conjugated to Alexa Fluor 647 (Invitrogen), and analyzed using a Becton Dickinson FACSCaliber flow cytometer.

Cryo-electron microscopy image reconstruction. DENV2 New Guinea C strain was mixed with Fab 1AID-2 at a concentration of one molecule of Fab for every E protein on the surface of the virus. The complex was incubated for 30 min either at room temperature or at 37 °C, followed by 2 h at 4 °C. Nearly all the particles incubated at 37 °C retained their ‘spiky-rough’ appearance after the temperature had been lowered to 4 °C, showing that the antibody was still bound. In contrast, only about one-third of the particles had a spiky-rough appearance when incubated at room temperature and then kept at 4 °C. Thus, the presence of antibody complexed with the virus at 37 °C on most of the particles showed that only a few or no particles had reverted to their original conformation when the temperature was lowered back to 4 °C. Precipitation at 4 °C before freezing might have improved the homogeneity of the sample by reducing the ‘breathing’ motion.

The virus Fab complex suspension was flash frozen on holey carbon grids in liquid ethane. Micrographs of the frozen complex were made with a CM200 FEG transmission electron microscope (Philips) using a calibrated magnification of 51,040 and an electron dose of approximately 25 e− per Å². The micrographs were digitized with a Zeiss SCAI scanner with a 2.74-Å separation between pixels. Particles were boxed and normalized using the program EMAN44. The micrographs were underfocuses by 2.8–3.6 µm Phases, but not amplitudes, were corrected by an appropriate calculation of the contrast transfer function. Uncomplexed West Nile virus was used as a starting model for image reconstruction. The orientations of the particles were determined using the program SPIDER45. The three-dimensional electron density map was calculated using the program XMIPP46, which had been modified to handle icosahedral symmetry. Out of a total of 2,885 boxed particles, 2,186 were used for the reconstruction of the 37 °C data to achieve a resolution of 24 Å, determined by dividing the particles into two equal sets and noting the resolution at which the Fourier shell correlation coefficient fell below 0.5 (Fig. 3a and Supplementary Table 2a, data set 3). The two leaflets of the lipid bilayers in the viral membrane were not resolved from each other, suggesting some positive staining in the sample.

To verify the validity of the reconstruction, the cryoEM density map generated from 532 particles using the sample incubated at room temperature (Supplementary Table 2a, data set 1) was compared with a reconstruction using a 540-particle subset of the 37 °C data (Supplementary Table 2a, data set 2). These two maps showed similar density distributions corresponding to a Fourier shell correlation greater than 0.5 at a resolution of 38 Å, similar to the resolution for both ~500 particle data sets (Supplementary Table 2a).

Accession codes. Coordinates for crystal forms 1 and 2 of the Fab 1AID2: E-DIII complexes have the PDB accession codes 2R69 and 2R29, respectively. The cryoEM map of the Fab complexed with native DENV has been deposited with The European Bioinformatics Institute and has the accession code EMD-1418. The coordinate for the fitted Fab and E molecules into the cryoEM map has the PDB accession code 2R6P.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank K. Choi for help with the crystallography, W. Zhang for advice concerning the cryoEM reconstruction and P.G. Lehman for help in data collection. We also thank Argonne National Laboratory and the Advanced Photon Source (Argonne, Illinois, USA) staff at beamline GM/C A 23 for help and encouragement with diffraction data collection. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11377. The GM/C A CAT is funded by the US National Institute of General Medical Science (Y1-GM-1104) and the US National Cancer Institute (Y1-CO-1020). We thank S. Johnson for help with the sequencing of the 1AID-2 V₃ and V₄ domains. The work was supported by the Pediatric Dengue Vaccine Initiative (TR-17) to R.J.K. and M.G.R. and a US National Institutes of Health (NIH) National Institute of Allergy and Infectious Diseases program project grant (AI053672) awarded to R.J.K., M.G.R. and others, M.S.D. and D.H.F. were supported by the Midwest Regional Center of Excellence for Biodefense and Emerging Infectious Disease Research (V541A057146), R.J.K. and M.G.R. also had support from Region V Great Lakes Regional Center of Excellence (NIH award U54 AI057153).
The Three-dimensional Structure of Genomic RNA in Bacteriophage MS2: Implications for Assembly

Katerina Toropova¹, Gabriella Basnak¹, Reidun Twarock²
Peter G. Stockley¹* and Neil A. Ranson¹*

¹Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK
²Departments of Biology and Mathematics, University of York, York YO10 5DD, UK

Received 29 June 2007; accepted 23 August 2007
Available online 7 September 2007

Edited by J. Doudna

Using cryo-electron microscopy, single particle image processing and three-dimensional reconstruction with icosahedral averaging, we have determined the three-dimensional solution structure of bacteriophage MS2 capsids reassembled from recombinant protein in the presence of short oligonucleotides. We have also significantly extended the resolution of the previously reported structure of the wild-type MS2 virion. The structures of recombinant MS2 capsids reveal clear density for bound RNA beneath the coat protein binding sites on the inner surface of the $T=3$ MS2 capsid, and show that a short extension of the minimal assembly initiation sequence that promotes an increase in the efficiency of assembly, interacts with the protein capsid forming a network of bound RNA. The structure of the wild-type MS2 virion at $\sim 9$ Å resolution reveals icosahedrally ordered density encompassing $\sim 90\%$ of the single-stranded RNA genome. The genome in the wild-type virion is arranged as two concentric shells of density, connected along the 5-fold symmetry axes of the particle. This novel RNA fold provides new constraints for models of viral assembly.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: cryo-electron microscopy; MS2; ssRNA virus; genomic RNA structure; atomic structure fitting

Introduction

All viruses face a common challenge of packaging their nucleic acids into a protective shell of protein subunits. Different viruses have evolved very different strategies to permit packaging of either DNA or RNA genomes whose functions at various stages of the viral lifecycle require them to adopt radically different conformations. The stratagem that many single-stranded (ss)RNA viruses have evolved is to assemble their capsids around their genomes. However, a major unresolved problem in structural virology is to understand the detailed molecular mechanism(s) that gives rise to the assembly of capsids of the correct size and symmetry. This is especially true for viruses whose capsids exhibit quasi-equivalent symmetry in their coat protein lattices.¹ Various proposals for formation of protein shells based on the initial formation of 3-fold,² or 5-fold³ assembly initiation complexes have been made previously. Capsid assembly is, however, spontaneous and usually very rapid, with the result that it has been technically difficult to isolate and characterise intermediates on the pathway to the final products beyond these "initiation complexes".⁴,⁵

The RNA bacteriophage MS2 is an ideal model for investigating such phenomena owing to the extensive biochemical and structural information that is available.⁶–¹¹ We have therefore undertaken a structural study of MS2 in order to address the specific question of the role(s) if any of the ssRNA genome during assembly. MS2 is a member of the Leviviridae family of viruses that infect male Escherichia coli cells via an initial attachment to the bacterial F-pilus. It has a single-stranded, positive-sense RNA genome of 3569 nucleotides that encodes just four gene products: coat protein (CP), replicase, lysis and maturation protein. CP is the most highly expressed of these four gene products and 180 copies assemble to form a $T=3$ icosahedral protein shell that encapsidates the genome in the mature
und is also incorporated into the virion and functions by binding the F-pilus. The maturation protein–genomic RNA complex is the only viral component to enter host cells during infection.\textsuperscript{12}

The crystal structure of the wild-type MS2 bacteriophage has been determined to 2.8 Å resolution.\textsuperscript{6,13} The CP fold was unique amongst known icosahedral viruses at that time, although it has since been shown to be shared by the other \textit{Leviviridae} coat proteins.\textsuperscript{10,14} The main-chain folds into a five-stranded antiparallel β-sheet with two antiparallel β-strands folding over it at the N terminus, and with a kinked α-helix at the C terminus (Figure 1). The C-terminal α-helices of two CP monomers interdigitate to form non-covalent dimers (CP\textsubscript{2}) in the capsid and these can be isolated as dimers in solution by acid dissociation.\textsuperscript{7,15} In the $T=3$ capsid, the CP is found in three distinct conformations, termed A, B and C, consistent with the quasi-equivalence symmetry required to construct a $T=3$ structure. The main site of variation between conformers is in the loop between the F and G β-strands, which is extended in A and C conformers, but bent back towards the main body of the subunit in the B conformer (Figure 1(a) and (b)). The capsid thus contains two types of dimer; A/B and C/C and the positioning of these different building blocks within the protein shell controls the size and symmetry of the viral particle.

Capsid re-assembly can be triggered \textit{in vitro} by a sequence-specific RNA–protein interaction between coat protein dimers and an RNA stem–loop (TR) of just 19 nt derived from the genomic sequence that encompasses the start codon of the viral replicase.\textsuperscript{16,17} RNA–protein binding thus achieves two functions \textit{via} a single molecular recognition event: translational repression of replicase and creation of an assembly competent complex on viral RNA. We have shown recently by NMR that TR-binding to CP\textsubscript{2} causes an allosteric conformational change within the protein from a largely symmetrical structure to an asymmetric one. Using mass spectrometry and size exclusion chromatography, it was also possible to show that the TR:CP\textsubscript{2} complex is kinetically trapped and forms capsid shells only very slowly. In contrast, when this complex is added to excess RNA-free protein, $T=3$ shells form rapidly, implying that the differing dimer conformers identified by NMR represent A/B and C/C-like species. Under these conditions the unit of capsid growth is a coat protein dimer and the intermediates in the assembly pathway are dominated by the formation of a species encompassing the first 3-fold axis of the particle.\textsuperscript{18} Formation of the 3-fold axis with its interdigitated A/B and C/C coat protein dimers commits the phage to assemble into a $T=3$ shell.

This is one of the most detailed molecular models for the switching of quasi-equivalent conformations available for any viral system. Allosteric, RNA-induced conformational switching, however, raises the question of how this effect is propagated during packaging of genomic RNA in which there is only a

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Figure_1.png}
\caption{Quasi-equivalence in the capsid of bacteriophage MS2. The structures of the MS2 coat protein dimer (CP\textsubscript{2}) in the (a) A/B and (b) C/C conformations are shown. The dimers are viewed from the outside, looking towards the centre of the particle, and the different conformers are coloured blue (A), green (B) and red (C). The main difference between conformers is in the orientation of the FG-loops (labelled). (c) The $T=3$ icosahedral assembly of 60 copies of the A/B dimer and 30 copies of the C/C dimer. The bent conformation of the FG-loop in the B conformer permits the A/B dimer to fit around the icosahedral 5-fold axes, whilst extended FG-loops of the A and C conformers interdigitate around the icosahedral 3-fold axes. All panels for Figures 1 and 2 were produced using PyMOL [www.pymol.org].}
\end{figure}
single occurrence of the TR sequence. We have shown that the kinetics of assembly are influenced by short extensions of the TR sequence with the natural genomic sequence, and that these effects are sensitive to the polarity of the added sequences.\textsuperscript{19} It is therefore important to understand the structure of the genomic RNA within the viral particle and the influence of sequences outside of the TR stem–loop on the assembly pathway.

Here we used cryo-electron microscopy (cryo-EM), single particle image processing and 3-D reconstruction with isosahedral averaging to determine the solution structures of $T=3$ shells reassembled with short genomic RNA fragments, and extend significantly the resolution of a previously reported solution structure for the wild-type bacteriophage MS2.\textsuperscript{20} The resulting structure for the MS2 virion, at a resolution of $\sim 9$ Å, provides novel insights into the organization of genomic RNA. The structure reveals a $T=3$ icosahedral protein shell within which isosahedrally ordered density for the majority of the genomic RNA is observed in an unprecedented structure consisting of two discrete but connected shells of density. The structures of the complexes with genomic RNA fragments are consistent with the idea that genomic RNA follows a defined path with respect to the coat protein shell and can participate actively in the assembly pathway.

Results

The structure of recombinant MS2 capsids reassembled with short genomic RNA fragments

We, and others,\textsuperscript{16,17} have shown that in vitro reassembly of $T=3$ MS2 capsids can be triggered by incubation of coat protein dimers with the TR 19 nucleotide stem–loop.\textsuperscript{16,17} We have also demonstrated that a 5’ extension of TR with 12 nt of genomic sequence (oligonucleotide S2) results in altered kinetics and yield of capsid formation in in vitro reassembly assays. This result suggests that sequences outside the minimal assembly initiation stem–loop play a role(s) in the assembly pathway.\textsuperscript{19}

To test this idea, we carried out reassembly reactions with both TR and S2 RNA oligonucleotides at an RNA:coat protein dimer molar ratio of 1:2 under conditions known to favour efficient production of the $T=3$ shell in vitro.\textsuperscript{18} The reassembly reactions were allowed to proceed for 3 h and then the resultant products were vitrified and examined in the electron microscope. Intact, reassembled capsids are the predominant species in solution after this period and comprise a relatively homogeneous population of particles that are readily identified and selected for image processing and structure refinement.

The starting model for our reconstructions was the crystal structure of the MS2 capsid lacking any coordinates for RNA,\textsuperscript{13} with all high-resolution features removed by application of a low-pass Fermi filter in Fourier space. The inside of such a starting model is therefore smooth and relatively featureless, with no protrusions of density into the encapsidated space (Figure 2(a) and (b)). The refined structure of the MS2 capsid assembled in the presence of TR RNA (TR–MS2) shows a strikingly different picture. Figure 2(c)–(e) shows the solution structure of TR–MS2 at $\sim 16$ Å resolution. It can readily be seen that significant extra density is present in the map compared to Figure 2(a) and (b). This density protrudes from the inside surface of the capsid towards the centre of the particle. In order to understand the positions of these densities in relation to the overlying lattice of CP, we docked the atomic coordinates for the MS2 capsid into our cryo-EM density. This (Figure 2(e)) clearly confirms that the extra densities are located beneath the CP sites where the TR operator is known to bind, i.e. across the ten-stranded antiparallel β-sheet. We therefore attribute these densities to bound TR RNAs. Owing to the isosahedral averaging used to calculate the reconstruction, all CP binding sites are occupied in the density map. The relative strength/significance of the protein and RNA electron densities, together with the biochemical observation that CP binds TR with high affinity suggest that the majority of such sites are occupied in solution. This is surprising given the stoichiometry of the assembly reaction, but it suggests that assembling intermediates bind stem–loops with higher avidity than coat protein dimers. This is consistent with the observation that TR stem–loops soaked into crystals of $T=3$ recombinant shells that do not contain tightly bound RNA fragments, bind to both A/B and C/C dimers.\textsuperscript{10,21}

The averaged electron densities below the A/B and C/C dimers are different and two modes of binding are expected based on the crystallographic data. At the symmetric C/C dimer TR RNA binds in two symmetry-related orientations.\textsuperscript{10} Within an A/B dimer, however, the conformation of the FG-loop in the B subunit appears to restrict the binding of the TR RNA so that at these sites it is found in a single orientation.\textsuperscript{10,13} In our isosahedrally averaged reconstruction, an averaging of the two conformations occurs resulting in a larger density beneath the C/C dimer compared to that found bound to A/B dimers.

The close similarity of our EM density to published crystallographic structures indicates that cryo-EM reconstructions of reassembled MS2 capsids accurately reflect their structure. We therefore have confidence that the reconstruction with the longer S2 RNA reveals the path taken by the additional genomic RNA sequence in the wild-type phage particle. The solution structure of S2–MS2 at $\sim 18$ Å resolution is shown in Figure 2(f)–(h). As expected, the TR and S2–MS2 reconstructions are essentially identical in the protein capsid region. The densities corresponding to bound RNA in the S2–MS2 map are larger, and encompass the densities observed in the TR structure, confirming the
Figure 2. Solution structures of MS2 capsids containing TR and S2 RNAs. The rear half (a) and a 30 Å thick central section through the density (b), of the RNA-free MS2 capsid crystal structure, converted to density at ~20 Å resolution. The RNA-free atomic coordinates have a smooth inner surface with no significant protrusions of density into the centre of the particle. (c) The sequences of the TR and S2 RNA oligonucleotides. The same views as shown for the RNA-free MS2 capsid ((a) and (b)) are shown for reassembled TR–MS2 ((d) and (e); pink) and S2–MS2 ((g) and (h); blue-grey). Reassembled RNA-containing capsids have distinct protrusions of density from the inner surfaces of the capsids, corresponding to bound RNA. (f) and (i) Difference density maps for TR–MS2 and S2–MS2 were generated by subtraction of the RNA-free crystallographic density from the respective cryo-EM maps. The density for RNA lying below the A/B and C/C dimers is shown (inset) for both TR–MS2 and S2–MS2. The crystallographic coordinates are shown in cartoon representation and coloured as for Figure 1. The resolution of the TR–MS2 and S2–MS2 maps was 16 Å and 18 Å, respectively, at a Fourier shell correlation (FSC) of 0.5 (see Supplementary Data).

interpretation that they represent RNA. The S2–MS2 map shows that the 5′ extensions of RNA interact with the protein lattice, forming the beginnings of a network of bound RNA, rather than extending into the centre of the particle and being smeared out by disorder and symmetry averaging. This observation is consistent with the network of genomic RNA found beneath the protein capsid for wild-type MS2 phage and other members of the Leviviridae. It is also consistent with the idea that these genomic sequences can play an active role(s) during the assembly process.

In both reconstructions the centres of the particles appear free of density. In this in vitro system where the only nucleic acids present are the oligonucleotides added, no artefactual density from symmetry averaging of noise was observed except at extremely low σ values, where disconnected noise appeared. This noise could be selectively removed using a low-pass Fourier filter. Indeed this was also the case in a further control where reconstructions were calculated from a test dataset derived from re-projections of a low-pass filtered, RNA-free crystal structure to which Gaussian noise was computationally added (not shown). The result with S2 oligonucleotides therefore suggests that the 5′ extended sequences as well as the minimal TR sequence interact with coat protein in the capsid, as expected from the earlier biochemical experiments.

The structure of wild-type bacteriophage MS2

Given the above result for recombinant MS2 capsids assembled with S2 RNA, it seemed likely that higher resolution information for the wild-type MS2 virion would be even more informative about the path of the genomic RNA across the inner surface of the protein shell. We therefore performed a new reconstruction of wild-type MS2 virions which yielded an electron density map at ~9 Å resolution. This structure, which represents a dramatic improvement in the available resolution for
the intact virus in solution, was calculated from 9335 images, which together with the icosahedral symmetry used for 3-D reconstruction equates to the averaging of ~560,000 asymmetric units. Consistent with previous X-ray crystallographic and lower-resolution cryo-EM studies, the reconstruction contains clearly resolved density for the T = 3 icosahedral coat protein lattice, with large pores through the shell at both the 3 and 5-fold icosahedral symmetry axes (Figure 3). To aid in the interpretation of this structure, the atomic coordinates for the MS2 capsid complexed with the TR oligonucleotide were docked into the EM density using the fitting software SITUS.23 The atomic coordinates fit the cryo-EM density extremely well, although the tips of the loops joining the first two β-strands (the AB-loop) in the A (shown in blue) and C (shown in red) conformers protrude somewhat from the cryo-EM map around the 3-fold axes. In previous cryo-EM reconstructions (including the TR– and S2–MS2 structures described above) density for this loop was not seen, yielding relatively smooth reconstructions which contrast starkly with available crystallographic structures where these loops project from the surface of the capsid giving MS2 a characteristic “spiky” appearance. The crystallographic B-factors for this loop are significantly higher than for the body of the protein (with the exception of the FG-loop), and are somewhat higher for this region in conformers A and C than in conformer B.13 This suggests both that the AB-loop is flexible and that it is somewhat more ordered in the B conformer. In our reconstruction we see these differences reflected in the cryo-EM map, with the loops largely out of the density for coat proteins in the A and C conformations, whereas in the B conformers only the very tips of the AB-loops protrude from a pronounced surface feature (Figure 3(e)). This implies that the loops are more ordered in the B quasi-equivalent subunits than in A and C, a context dependent consequence of quasi-equivalence.

The organisation of genomic RNA in WT-MS2

Whilst the cryo-EM structure of the MS2 protein capsid itself is consistent with previous studies, when the interior of the virion is examined entirely new features are observed. Figure 4 shows difference density maps where the cryo-EM density corresponding to the capsid region has been selectively removed, leaving only the density features within the interior of the capsid. Sections through this masked WT–MS2 reconstruction, together with the atomic coordinates for the capsid, are shown perpendicular to the icosahedral 2-fold, 3-fold and 5-fold axes (Figure 4(a)–(c), respectively). Substantial ordered density, representing the encapsidated genomic RNA occupies the interior of the virion as two connected shells (see Supplementary Data, movie 1). The outermost shell lies directly beneath the protein capsid between (averaged) radii of ~108 Å and ~84 Å. In agreement with previous EM26 and crystallographic studies with short RNA

Figure 3. The solution structure of the wild-type bacteriophage MS2 virion. (a) Surface view and atomic structure fitting of the WT–MS2 cryo-EM structure. On the left, the cryo-EM density envelope is shown as a solid grey surface. On the right, the fitted atomic coordinates ([PDB ID: 1aq3]) are included in a cartoon representation (coloured as for Figure 1), embedded within a transparent surface. The resolution of the map was 8.9 Å (FSC = 0.5; see Supplementary Data). (c)–(e) Details of the fit of atomic coordinates into the WT–MS2 structure. The views shown are perpendicular to the (b) 2-fold, (c) 3-fold and (d) 5-fold global symmetry axes. (e) The different quasi-equivalent environments for the A, B and C coat protein conformers. The upper portions of the AB-loops in conformers A and C do not fit into the EM density, whilst for conformer B, all but the very tip of this loop is buried in a pronounced surface feature (in the foreground of the Figure). All images in Figures 3–5 were created using UCSF Chimera [http://www.cgl.ucsf.edu/chimera].
fragments, this shell consists of a connected network of density that maps the binding sites for RNA on the inner surface of the coat protein lattice. The second density shell is found at much lower radii, occupying a region extending from ∼42-56 Å from the centre of the virion.

Around the icosahedral 5-fold axis, weak connections are made from the protein-bound RNA layer, to a substantial density that extends down the 5-fold axis for more than 40 Å towards the centre of the virion. These densities run from the edge of the outer RNA shell (r ≈ 82 Å) to the innermost edge of the inner shell (r ≈ 40 Å). The volume of the major densities lying along the 5-fold axes appears only large enough to encompass a single duplex of RNA. At low radii (40-50 Å), further connections between the densities on the 5-fold axes are seen. At radii below the innermost edge of the inner shell, the centre of the virion is empty.

There are no features in the icosahedrally-averaged cryo-EM density map that can be assigned to the single copy of the maturation protein present in the virion. Since this protein must be located at least in part at the surface of the phage particle, it is possible that some of the density along the 5-fold axis arises from inappropriate icosahedral averaging of this protein.

The combined volume of the density that we ascribe to RNA accounts for ∼25-30% of the internal volume of the capsid. Owing to the wide range of published estimates for the density of RNA, the fraction of the genomic material encompassed by this density is difficult to estimate accurately. However, using the volume occupied by the nucleotides of TR seen in crystallographic studies as a guide, we estimate that the encapsidated density we attribute to RNA may accommodate ∼90% of the 3569 nucleotides in the wild-type MS2 genome. Given the significant and obvious errors implicit in such a calculation, there remains the possibility that the entire MS2 genome may be packaged with near-icosahedral order at this resolution, an unprecedented degree of ordering for a single-stranded RNA virus.

**Binding of genomic RNA to the inner surface of the WT-MS2 capsid**

Figure 5 shows details of the interaction between the genomic RNA density and the inner surface of the WT-MS2 capsid. The density in this RNA shell is the result of icosahedral symmetry averaging. Nevertheless the RNA density below each type of quasi-equivalent coat protein dimer is different. For dimers in the A/B conformation, density for bound RNA is observed beneath the dimer interface and extends towards the icosahedral 5-fold axes. These axes are surrounded by dimers in the A/B conformation, and so a connected ring of density is formed encircling each 5-fold axis. Weak density connections are made from this ring to the inner shell of RNA. The RNA density beneath the C/C dimers also sits beneath the dimer interface, but the connectivity is different to that at the A/B site (Figure 5), suggesting that the average interaction between coat protein and genomic RNA is different at these positions. This is precisely what the allosterically mediated model of the assembly pathway predicts.

**Discussion**

The range of different solutions that viruses have evolved to overcome the challenge of packaging their genomic material is very broad. Double-stranded (ds)DNA viruses use powerful packaging motors that actively insert their base-paired genomes into pre-formed pro-capsids. Such active packing typically results in a close-packed, spooled array of DNA duplexes with little if any free volume remaining within the viral capsid. At the other extreme, many RNA viruses assemble their capsids spontaneously, coincident with folding of their
single-stranded genomes. However, the mechanisms by which capsid assembly and genome folding operate and interact remain poorly understood.

The structure of the MS2 genome in the viral capsid

The reconstruction of the wild-type MS2 phage particle at intermediate resolution (≈9 Å) presented here gives a unique insight into the structure and organization of the viral genome. In contrast to the dsDNA viruses discussed above, the MS2 virion is remarkably empty. Only approximately one-quarter of the encapsitated volume of the bacteriophage is occupied with density for RNA, which together with the substantial pores through the capsid structure at both the 3- and 5-fold symmetry axes, suggests that those portions of the folded RNA molecule not bound to protein will be fully hydrated.

X-ray crystallography of the same bacteriophage at 2.8 Å resolution showed no density for the RNA component other than weak, disconnected density ascribed to a purine below the known TR adenine (A-4) binding pocket residues. This difference suggests that the icosahedral ordering seen in the cryo-EM density does not extend to high resolution. The relative electron density of the RNA in the two inter-connected shells is similar in some areas to the levels observed for the coat protein shell, suggesting that these features in the RNA reflect relatively full occupancy and do not simply represent an average of several differing conformations within the capsid. In contrast there is a clear weakening of the density of the connectors as they extend towards the outer RNA shell. This is consistent with averaging around the overlying 5-fold axis in the particle, whilst the major connections to the inner shell remain strong. The implication is that RNA is transiting between shells from different points around the outer 5-fold axis but ending up in a unique position on the inner shell. Such an inter-connected pair of shells of genomic RNA density has not been observed previously, although this type of organization may be fairly common. For instance there is a well-ordered dodecahedral cage of RNA located underneath the protein shell of Faricato virus, but this only accounts for some 35% of the genomic RNA, the remainder presumably being located at lower radii but not visible in either X-ray or cryo-EM maps.

The high degree of ordering of the MS2 genome at this resolution is of interest because unlike many other ssRNA viruses the MS2 coat protein lacks N-terminal extensions (or “arms”) that can interdigitate with the packaged genome and thus propagate the icosahedral symmetry of the protein shell throughout the particle. In the case of MS2 the ordered structure observed must arise principally as a consequence of RNA folding. There are, however, several sequence-specific contacts to protein components of the virion that must also help to position defined RNA sequences within this fold(s). These are the assembly initiation complex made by the coat protein with the TR stem-loop and the contacts to sequences located towards the genomic 5’ and 3’ ends that are made by the single copy of maturation protein. Given the locations of these proteins, i.e. either wholly (CP) or at least partially (maturation protein) on the outer surface, it is tempting to speculate that all three of these RNA sequences must...
be in the outer shell of RNA. Alternatively if the maturation protein were spatially extended it could span all three major molecular layers of the virion making one or both contacts with the RNA in either shell. These three specific constraints, as well as the network of non-sequence specific interactions required to switch quasi-equivalent conformers, and the need to fold into the double shell structure with defined size and symmetry, suggest that the observed cryo-EM density corresponds to a relatively small, well-ordered ensemble of permisssible RNA folds. This proposal is inherently testable and appropriate experiments are underway to examine it.

Implications for MS2 assembly

The structures of the reassembled particles in the presence of short genomic sequences are consistent with the inferences made on the basis of biochemical reassembly assays, namely that genomic sequences outside the TR stem–loop can influence the assembly pathway via interaction with CP subunits.\textsuperscript{18,19} We have shown in TR-induced reassembly reaction that the TR-bound CP\textsubscript{2} is asymmetric and the complex metastable until excess RNA-free CP\textsubscript{2}, which is primarily a symmetric dimer, is added.\textsuperscript{18} Once both types of dimer are present assembly of the $T=3$ shell is rapid and proceeds by coat protein dimer addition, apparently through intermediates that are dominated by stoichiometries corresponding to the formation of the particle 3-fold axis, i.e. $3 (TR+CP_2)+3(CP_2)$. The clear implications are that the binding of TR RNA converts a CP\textsubscript{2} from a C/C-like species to an A/B-like species and that both are required for efficient assembly. This RNA-induced, allosteric switching between quasi-equivalent conformers completely explains assembly in the presence of multiple copies of the TR sequence. Similar assays in the presence of ssRNA show significant increases in the rate and yield of assembly.\textsuperscript{19} These can now be understood in structural terms, since the additional nucleotides interact with a neighbouring coat protein dimer. It is interesting to speculate whether such interactions are completely indifferent to the sequence of this RNA extension.

The assembly pathway followed when genomic RNA is being packaged cannot, however, be identical to that seen with the TR oligonucleotide. Genomic RNA sequences play essential roles in virion assembly, and the RNA molecule must be compacted to fit within the mature virion. It must also serve as the message for viral protein synthesis, and as a substrate for RNA replication. The kinetics and mechanism of folding this RNA into the two-shell structure are obviously major factors in compacting the genome to the dimensions of the eventual $T=3$ shell. Whether the RNA folding occurs before or during capsid assembly is currently unknown, although the latter mechanism would be akin to the assembly of the ribosome, another complex containing both protein and large, complex and highly folded RNA molecules. What is clear from the cryo-EM reconstruction is that networks of RNA–protein contacts are created throughout the virion that are, on average, distinct at A/B and C/C dimers. This observation is consistent with the idea that sequences/secondary structures in the genome other than TR can contribute to quasi-equivalent conformer switching at defined positions throughout the assembly process. Presumably this genome-mediated assembly pathway is much more efficient than reassembly in the presence of short fragments in sub-stoichiometric amounts which is also known to occur.

Such an RNA-templated capsid assembly pathway has a number of features that constrain the RNA folding even further. Larson \& McPherson\textsuperscript{50} have addressed this question for satellite tobacco mosaic virus (STMV), a $T=1$ structure where up to 80\% of the genomic ssRNA can be accounted for as icosahedrally ordered stem–loops located in a single shell complexed with the coat protein. They argued that the STMV RNA packaged in the virion is likely to form secondary structures dominated by local stem–loop interactions formed as the genomic transcript emerges from a replication complex. Such local structural motifs are separated by single-stranded regions that allow the RNA a great deal of conformational freedom, permitting local sequence/structure variations to occur whilst still fulfilling the need to make interactions with each facet of the protein lattice. In this model the packaged RNA structure need not correspond to the lowest free energy or physiologically active forms of the molecule.

The situation must be different in MS2 for two reasons. Firstly we have successfully reassembled $T=3$ shells from \textit{in vitro} transcribed and purified large sub-genomic fragments of RNA that have had the opportunity to form long range secondary and tertiary interactions (unpublished data). Secondly the formation of a double-shell structure of RNA imposes additional constraints on the path of the RNA within the virion. A useful way to think about this problem is to project the outer RNA network onto a plane net of the protein shell (Figure 6).\textsuperscript{1,16} Figure 6(a) shows the locations of the capsid protein dimers relative to the surface of an icosahedron in a planar representation. This is a schematic view of the spherical arrangement in Figure 6(b) and facilitates comparison with the location of the RNA density (Figure 6(c)). The RNA within the outer shell forms a polyhedral cage (Figure 6(d)) with 60 vertices located underneath the 60 A/B dimers. The contacts between RNA and A/B dimers are important for conformational switching from a symmetric to the required asymmetric form, implying that the RNA meets every vertex of this polyhedral cage. In contrast, this model does not require the path of the RNA to traverse all the edges of the polyhedron, which are positioned underneath the C/C dimers. The RNA density observed beneath C/C dimers could result from secondary structure elements arising at either vertex bordering that edge.

We also know that capsid assembly takes place by addition of coat protein dimers to the growing
intermediates, with A/B dimers resulting from binding to (stem–loops of) RNA. Due to the requirement to make these RNA–protein interactions, the formation of the folded RNA shell and of the protein lattice are likely to occur simultaneously. Folding of the RNA into secondary and tertiary structures appropriate for packaging, if necessary, may be facilitated in part by the interaction with coat protein. As Figure 6(c) shows, even if assembly is initiated at a single point (TR) on the RNA there are multiple sub-pathways that intermediates could then follow to reach the final \(T=3\) capsid.

There could be further constraints implied by the connections between the inner and outer RNA cages. As shown in Figure 4, the two shells of RNA density are connected along the particle’s 5-fold axes. The observed density of these connections is consistent with the presence of at most a single RNA duplex or perhaps two single strands of RNA (Figure 7(a)) at every 5-fold axis. If we consider the path of the genomic RNA around the outer shell, at some point it must cross to the inner shell and then connect back to the outer shell. The two possible ways in which these transitions could occur are shown in Figure 7(b) and (c). In the first of these, the RNA strand leaves the outer shell, and returns to it, at the same 5-fold axis (Figure 7(b)). The alternative is that RNA strand leaves at one 5-fold axis and returns at a different 5-fold axis, i.e. that it leaves at one vertex of the outer shell and returns to a neighbouring position (Figure 7(c)). If the latter were correct, a single strand RNA cannot account for the observed density, so a second strand would be required to transit in the opposite direction at the original 5-fold axis after following a distinct pathway. Moreover, the single RNA strand would first have to transit back to the outer shell at a different 5-fold axis, with no obvious constraints on the location of that axis. This seems highly unlikely given the efficiency of capsid assembly and the propensity of RNA to form base-paired duplexes. If the former were true, the connections and the inner shell could be composed of discrete secondary structural elements, thus imposing far fewer constraints on RNA folding and evolution.

**Implications for phage infection**

RNA phages associate with target *E. coli* cells by binding to the F pilus via the maturation protein.\(^{12}\) Only the maturation protein–genomic RNA complex then enters the cell via a mechanism that is currently not understood. The cryo-EM structure of the MS2 virion presented here strongly suggests that, in sharp contrast to double-stranded (ds)DNA viruses, there is little or no driving force within the
capsid to achieve this genome release. The RNA genome is not close-packed within the virion, and the capsid is porous to small molecules. This in turn suggests that the driving force for genome release must be derived from some function of the host bacterium. A recent low-angle neutron scattering study suggested that the dimensions of the MS2 capsid proteins are different in the presence or absence of maturation protein, the form lacking maturation protein being considerably thicker (31–37 Å versus 21–25 Å) than in wild-type particles. This difference was ascribed to the formation of a post-infection state within the capsid lattice when both RNA and maturation protein have left the protein shell. This state was not revealed by previous X-ray structures, even those of recombinant shells lacking maturation protein, it was argued because crystallization favours the “thin” pre-infection state. We see no evidence for two such states in our data. The solution structures of both reassembled and wild-type phage map onto the X-ray structures of both phage and recombinant protein shells with high fidelity.

Conclusions

High-resolution structures for large RNA molecules are rare, and the MS2 genomic RNA structure presented here is amongst the highest resolution for such a large RNA molecule (by comparison the 23S rRNA of the prokaryotic 50S large ribosomal subunit determined at atomic resolution is ~2900 nucleotides in length). Positive sense, single-stranded RNA viruses are the most common type of viruses in nature. Unlike many of these, however, MS2 does not interact with its genome via extensions (or “arms”) from the body of its coat protein. The data presented here provide the first direct evidence that a viral shell lacking such arms to propagate icosahedral symmetry through the particle, can promote the folding of a large RNA molecule into such a compact, ordered fold. Such large RNA molecules are relatively common and are of enormous importance in cells. It is known that many of their functions are regulated by RNA folding. The double-shell structure seen here may well represent a general architecture for compacting such large molecules into non-active states.

Materials and Methods

Proteins, RNAs and reagents

Wild-type bacteriophage MS2, maintained in buffer A (10 mM Tris–HCl (pH 7.5), 100 mM NaCl, 0.1 mM MgCl₂ and 0.01 mM EDTA), was a gift from Professor David S. Peabody, University of New Mexico (Albuquerque). Wild-type, recombinant MS2 coat protein (CP) was overexpressed in E. coli and purified using described methods. Purified T = 3 capsids were stored in 10 mM Hepes (pH 7.2), 100 mM NaCl and 1 mM EDTA. CP₂ for reassembly reactions was obtained by treating the recombinant capsids with glacial acetic acid to induce capsid dissociation to disassembled CP monomers, followed by exchange into pH 5.2 buffer to promote dimerisation. To ensure that complete disassembly had occurred before reassembly, the CP₂ were examined using negative stain EM. Synthetic oligonucleotides, TR and S₂, were prepared as described.

Recombinant MS2 reassembly reactions

Capsid reassembly was carried out by mixing TR or S₂ with CP₂ in a 1:2 molar ratio in buffer B(20 mM Tris-acetate, 0.4 mM magnesium acetate, 20 mM acetic acid (pH 5.2)) on ice. After reassembly for 3 h, samples were flash-frozen as described below.
Cryo-EM data collection and image pre-processing

The 3 μl aliquots of wild-type MS2 at 0.8 mg ml⁻¹ in buffer A and reassembly reaction samples in buffer B were placed onto 300-mesh Quantifoil R2/1 grids that were glow-discharged in air for ~30 s immediately before use. Grids were frozen by plunging into liquid nitrogen-cooled, liquid ethane, using a computer-controlled, pneumatically driven freezing apparatus.35 Samples were imaged on an FEI Tecnai-F20 microscope equipped with a Gatan 626 cryo-transfer stage, using low-dose protocols (~15 e⁻/Å²). Wild-type MS2 was imaged on Kodak SO-163 film at 50,000× magnification and micrographs were digitized using a Nikon Coolscan3000 scanner at a final object sampling of 1.25 Å/pixel. Reassembly reactions were imaged on a GATAN US4000SP, 16Mpixel CCD camera at 50,500× magnification, giving a final object sampling of 2.97 Å/pixel.

Micrograph defocus and astigmatism were determined computationally using the program CTFIND3,36 and micrographs showing significant drift or astigmatism were discarded. All particles recognizable as isolated MS2 capsids were semi-automatically selected using the program BOXER37 and correction for the microscope contrast transfer function was performed by computational filtering of image phases in SPIDER.38 All other image processing steps were performed in SPIDER unless otherwise stated. For wild-type MS2, 10,382 molecular views were extracted into 448 × 448 pixel boxes and band-pass filtered between 350 Å and 4 Å. For reassembly reactions with TR and S2 RNAs, 1550 and 1715 molecular views, respectively, were extracted into 320 × 320 pixel boxes and band-pass filtered between 350 Å and 6 Å. All extracted images were normalised to a constant mean and standard deviation.

Single particle reconstruction

The atomic coordinates for wild-type MS2 protein capsid (PDB ID: 2ms2)39 were converted to density and all high resolution features removed by Fourier filtration to 40 Å using a low-pass Fermi filter. This low-resolution, RNA-free representation of the MS2 capsid was then used as a starting model for EM structure refinement of all the data. For wild-type MS2, the low-resolution density was projected across 78 orientations to give a range of molecular views evenly covering the icosahedral asymmetric unit with a spacing of 3°. All image data were then iteratively aligned to these views. Following each alignment, averages of the images corresponding to each view were calculated. 3-D reconstructions were generated after each refinement round using weighted back-projection and imposing icosahedral symmetry. The resulting reconstruction was then used to generate a new series of reference views, after masking with a tight, soft-edged mask derived from the density map, and the whole procedure iterated. In the final rounds of refinement, images were ranked according to cross-correlation coefficient and the worst aligning members for each orientation were excluded. This was implemented in such a way that the best-populated molecular view contained no more than three times as many raw images as the worst populated view. The resolution of the final reconstruction, calculated from 9335 raw images, was determined to be 9 Å using the 0.5 Fourier shell correlation criterion (see Supplementary Data).

For reassembly reactions of recombinant MS2 capsids, 3-D reconstructions were generated using the method described above, with the exception that the low-resolution starting model was projected across 53 orientations (angular spacing of 3.5°). Final reconstructions contained 957 and 1152 individual molecular views for TR and S2, respectively, and the final resolutions were determined to be 15.5 Å and 17.8 Å, using the 0.5 Fourier shell correlation criterion (see Supplementary Data).

Masking and RNA volume estimation

To remove density corresponding to the protein shell from the WT-MS2 cryo-EM map, a mask was made by taking the crystallographic coordinates for the protein shell, converting to density and removing high-resolution features by Fourier filtration to the same resolution as the cryo-EM map. This density was then converted to binary and inverted to form a hard-edged mask. The edges of this mask were then softened with a second Fourier filtration step using a relatively sharp Fermi filter profile in SPIDER. Finally, the cryo-EM maps were multiplied by the mask to remove protein density.

A benchmark figure for the per-base volume of RNA was estimated by taking the 60 copies of the TR oligonucleotide present beneath A/B dimers in the crystal structure of the MS2–TR complex (1AQ3), which each contain 16 ordered nucleotides (960 in total), and converting to density in SPIDER. The resulting density was Fourier filtered to remove high-resolution information. The combined volume of these densities, and of the RNA density in the WT-MS2 structure were then calculated in UCSF Chimera.

Atomic structure fitting

An initial fit of the MS2 capsid was generated by manual docking of the atomic coordinates of the MS2 capsid (PDB ID: 1AQ3) into the wild-type MS2 EM density map using PyMOL.† The orientation of this manual fit was then refined using the program SITUS.35 This refined orientation was also used for the cryo-EM map for capsid reassembled with TR, which is in the same coordinate system owing to the image processing strategy employed. The cryo-EM maps for WT-MS2, TR-MS2 and S2-MS2 are deposited in the EMDB as EMD-1431, EMD-1432 and EMD-1433, respectively.

Acknowledgements

We thank Professor David Peabody, University of New Mexico for his gift of wild-type MS2 bacteriophage, Dr Sean Killen and Mike Walls for computing support, and Dr Abdul Rashid for oligonucleotide synthesis. We thank Ottar Rolfsson for discussion of unpublished data. K.T. is a research student supported by the University of Leeds Interdisciplinary Institute (UII) in Bionanosciences, which also funded the work. R.T. is an EPSRC Advanced Research Fellow and P.G.S. acknowledges financial support from the UK BBSRC and

The Leverhulme Trust. N.A.R. holds a University of Leeds Research Fellowship.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2007.08.067

References

Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction

Xing Zhang*, Ethan Settembre†, Chen Xu†, Philip R. Dormitzer†‡, Richard Bellamy§, Stephen C. Harrison†‡∗∗, and Niko La Grigorieff‡∗∗∗

*Department of Biochemistry and Howard Hughes Medical Institute, and †Rosenstiel Basic Medical Sciences Research Center, Brandeis University, MS029, 415 South Street, Waltham, MA 02454; ‡Laboratory of Molecular Medicine, and †Howard Hughes Medical Institute, Children’s Hospital, 320 Longwood Avenue, Boston, MA 02115; and §School of Biological Sciences, University of Auckland, Auckland, New Zealand

Contributed by Stephen C. Harrison, December 10, 2007 (sent for review November 26, 2007)

Electron cryomicroscopy (cryo-EM) yields images of macromolecular assemblies and their components, from which 3D structures can be determined, by using an image processing method commonly known as “single-particle reconstruction.” During the past two decades, this technique has become an important tool for 3D structure determination, but it generally has not been possible to determine atomic models. In principle, individual molecular images contain high-resolution information contaminated by a much higher level of noise. In practice, it has been unclear whether current averaging methods are adequate to extract this information from the background. We present here a reconstruction, obtained by using recently developed image processing methods, of the rotavirus inner capsid particle (“double-layer particle” or DLP) at a resolution suitable for interpretation by an atomic model. The result establishes single-particle reconstruction as a high-resolution technique. We show by direct comparison that the cryo-EM reconstruction of viral protein 6 (VP6) of the rotavirus DLP is similar in clarity to a 3.8-Å resolution map obtained from X-ray crystallography. At this resolution, most of the amino acid side chains produce recognizable density. The icosahedral symmetry of the particle was an important factor in achieving this resolution in the cryo-EM analysis, but as the size of recordable datasets increases, single-particle reconstruction also is likely to yield structures at comparable resolution from samples of much lower symmetry. This potential has broad implications for structural cell biology.

FREALIGN | virus structure | rotavirus | protein structure | image processing

Cryogenic preservation of macromolecular assemblies was introduced into electron microscopy 30 years ago (1, 2). Different applications of electron cryomicroscopy (cryo-EM) include 2D crystallography, helical reconstruction (helices are effectively rolled-up 2D lattices), single-particle reconstruction, and tomography (reviewed recently in ref. 3). In several cases, data collected from 2D crystals and helical specimens have led to structures at 4-Å resolution or better, enabling their interpretation by atomic models (for example, see refs. 4–6). The periodicity present in images of 2D crystals and helical specimens generates distinct diffraction patterns with signal-containing spots and lines that can be separated from the background. Removal of the background (noise) is straightforward and amounts to averaging over all unit cells in the crystal. The distinction between signal and background is not possible in images of single particles and tomographic reconstructions. Instead, in the case of single-particle reconstructions, averaging must be performed over many individual particle images. The resolution obtained by this averaging procedure depends critically on the accuracy with which each particle can be brought into register with the others. This computational step is not an issue for well ordered 2D crystals and helical specimens used in high-resolution studies because intermolecular contacts establish the alignment of molecules. Other important imaging parameters, such as the magnification, defocus, and beam tilt used to collect the images in the microscope, also are easier to determine for 2D crystals and helical specimens because a single image typically contains 100 to 1,000 times more signal than an image of an individual molecule or complex (3). For example, a purple membrane crystal containing ~140 x 140 unit cells has a molecular mass of ~1,760 MDa (7), generating ~700 times the signal of a bacterial ribosome of 2.5 MDa (8). Finally, in most cases, a carbon film is used to support the crystalline specimens, providing additional stability to the specimen that can reduce movement during exposure to the electron beam (9).

Because of the lack of periodicity and, in many cases, the lack of a stable support film, single-particle reconstruction techniques have so far fallen short of yielding structures interpretable by atomic models. Efforts to achieve “atomic resolution” have continued, however, because the single-particle technique is applicable to a wide range of noncrystalline samples and can be used with just a few picomoles of material (10). Furthermore, single-particle techniques also can be applied to individual unit cells or mosaic blocks of crystalline samples, where they can perform better than crystallographic and helical techniques if the sample is partially disordered (11, 12). The single-particle technique therefore assumes a central role in cryo-EM. The cryo-EM-based 3D reconstruction of an icosahedral virus particle described in this article has a resolution adequate for tracing a polypeptide chain. The result demonstrates directly that recently developed computational techniques (13) can extract atomic-level detail from images recorded with contemporary instrumentation.

Rotaviruses are multisubunit, nonenveloped, icosahedrally symmetric viruses with an 11-segment, double-strand RNA genome (14). Infectious virions are sometimes called “triple-layer particles” because three distinct protein shells surround the packaged viral RNA. During cell entry, the outer layer is lost, and the remaining “double-layer particle” (DLP), which contains approximately one copy of the viral RNA-dependent RNA polymerase and the capping enzyme for each genomic segment, generates mRNA transcripts and exports them into the cytoplasm of the infected cell. The inner layer of the DLP, which
encloses the RNA, contains 120 copies of viral protein 2 (VP2; 102 kDa); the outer layer contains 780 copies of VP6 (41 kDa). VP6 forms stable trimers, which pack outside the VP2 layer in a T = 13 icosahedral array. A recently completed x-ray crystal structure of the DLP has yielded a full atomic model, refined at 3.8-Å resolution (B. McLain, E.S., R.B., and S.C.H., unpublished data). A previously determined structure of isolated VP6 trimers provides an even more precise view of that component (15).

Results

Quality of Data. Data were collected from two DLP preparations. We aligned 18,125 images of rotavirus DLPs frozen in ice (Fig. 1) and merged a subset of these into a 3D reconstruction by using the computer program FREALIGN (13). Analysis of the aligned particles, which have a diameter of ≈710 Å, revealed a clear bimodal distribution of correlation coefficients between individual images and the refined reconstruction. A particle parameter search (see Materials and Methods) moved ≈13% of the particles from the lower cluster to the cluster with a higher correlation. This search was computationally expensive and had to be limited to a repeat of 50 times. The number of moved particles was found to depend on the number of search orientations, suggesting that one reason for a low correlation coefficient is misalignment. Other reasons may include more subtly damaged or otherwise disordered particles. An analysis of averaged power spectra calculated from particles selected from both clusters showed Thouless rings (16) to a resolution of ≈6 Å, suggesting that the image quality is comparable in both cases. Selecting only the 8,400 particles from the cluster with a higher correlation, we computed a reconstruction with a resolution of ≈5 Å (Figs. 2–4). The icosahedral symmetry of the particle was applied during reconstruction, increasing the effective size of the dataset 60-fold. We also used the remaining particles to calculate a second reconstruction (data not shown), which had a resolu-

Fig. 1. Image of rotavirus DLPs frozen in ice over a hole in C-flat carbon grids. The arrows indicate particles that differ significantly in appearance from other particles in the image, indicating partially damaged particles. These particles were not used for further processing. More subtle damage to particles that is not readily visible may be one reason for the lower correlation coefficients with the reference seen for many particles (Fig. 5).

Fig. 2. VP6 trimers in the DLP. (a) Rotavirus DLP structure filtered at 20 Å. VP6 trimers involved in the 13-fold nonicosahedral averaging are indicated by letters. The T trimer coincides with one of the icosahedral threefold axes and contains only one unique VP6 monomer. The other four trimers contain the additional 12 VP6 molecules that were used for nonicosahedral averaging. (b) T trimer of VP6, filtered at 7-Å resolution, with a bundle of three generic helices docked into easily identifiable density features; 12 other bundles were docked in equivalent density in the other four VP6 trimers. (c) Overview of the 13-fold averaged VP6 trimer at 3.8-Å resolution. Different areas are highlighted and labeled with numbers. They are shown in more detail in d and e and in Fig. 6. (d) Density in area 1 before 13-fold averaging, shown with the atomic model of VP6, determined by using the x-ray density map (B. McLain, E.S., R.B., and S.C.H., unpublished data). (e) Same as in d but after 13-fold averaging. All structures were prepared with UCSF CHIMERA (35).

Fig. 3. Density map of part of the VP2 portion of DLP. The map was filtered at 4.5-Å resolution and sharpened by using a B factor of −450 Å². The quality of the map is similar to that of VP6 before nonicosahedral averaging (Fig. 2d).
tion of only 25 Å. Thus, the correlation coefficient is a clear indicator of a usable particle.

The number of usable particles depends strongly on the batch number of the virus DLP preparation. Approximately 30% of the virus particles collected from the first batch fell below the threshold criterion for inclusion in the reconstruction, whereas it was ~75% in the second batch (Fig. 5a). Data from both batches were recorded on several days and from different cryo grids, and properly aligned particles from both batches had approximately the same average correlation coefficient. Almost all micrographs contained both usable and discarded particles. However, virus particles from the first batch were frozen by using homemade lacy carbon, whereas material from the second batch was frozen by using commercially available C-flat grids that were lightly recoated with carbon before use to enhance conductivity. The dependence on batch number might reflect a difference in the quality of the particles. It is more likely, however, that the data from the second batch contained more misaligned particles because they were added at a different stage in the image processing, compared with the first batch (see Materials and Methods).

The correlation coefficient also depends on the defocus used to collect the data (Fig. 5b). This dependence is not surprising because the defocus determines the spectral distribution of the signal in the image. A larger defocus amplifies the signal at low resolution where the molecular transform of the virus particle is strongest. Corresponding particles therefore show a higher correlation coefficient than particles imaged with a smaller defocus. The correlation coefficient also depends to some degree on the location of the virus particle in the carbon film hole (Fig. 5c). Particles closer to the edge of a hole tend to have a smaller correlation coefficient than particles closer to the center. We measured the ice thickness by burning a small hole into the ice layer at a tilt of ~45° and then observing the hole length at a tilt of +45°. The measurements showed that the ice thickness close to the carbon was ~1,200 Å, whereas it was only 700 Å near the center of the hole, similar to the diameter of the DLP. Thicker ice increases the fraction of inelastically and multiply scattered electrons, which add to the background in an image and reduce the signal produced by the remaining singly and elastically scattered electrons. This reduction in signal lowers the correlation coefficient between particle and reference. Some particles too close to the hole center also exhibited a slight decrease in correlation coefficient. These particles might have suffered some deformation because of the thin ice. The dependence on location was least noticeable in samples prepared by using C-flat grids, suggesting that the ice thickness in those samples was more uniform than in grids prepared with lacy carbon.

**Nonicosahedral Averaging.** We used the atomic models for VP2 and VP6 built into the x-ray map to interpret the cryo-EM structure. In the icosahedrally averaged reconstruction of the DLP, the density corresponding to most aromatic side chains is clearly visible, whereas other, mostly smaller, side chains appear in incomplete or displaced density (Figs. 2d and 3).

The 780 VP6 monomers in the outer shell of DLP assume T = 13 icosahedral symmetry. The density map of VP6 therefore could be improved further by averaging the density of the 13 equivalent monomers that are not icosahedrally related. Bundles of three generic α-helices, generated with the computer program MOLEMAN (17, 18) and containing one 8-alanine and two 11-alanine helices, were placed in clearly identifiable density features of the three VP6 monomers around one of the icosahedral threefold axes. For the placement, we used a density map filtered at 7-Å resolution and sharpened with a B factor of ~350 Å² (Fig. 2h). The low-pass filter had a cosine-shaped cut-off with a width of ~5 Fourier pixels. Copies of the trimer of helical bundles then were placed in the four other VP6 trimers densities containing the remaining 12 equivalent monomers (Fig. 2a). Approximate matrices transforming the coordinates of these four trimers into the trimer on the icosahedral threefold axis
were determined by using the computer program LSQMAN (17, 18). The matrices were refined with the computer program IMP (17, 18) by aligning densities in the vicinity of the trimers of helical bundles against equivalent density of the VP6 trimer on the threefold icosahedral symmetry axis. For the refinement, we used a density map filtered at 4-Å resolution and sharpened with a B factor of ~300 Å². As before, the low-pass filter had a cosine-shaped cut-off with a width of ~5 Fourier pixels. A full set of 12 transformation matrices was derived from the 4 initial matrices by combining the matrix transformations with additional 120° rotations around the threefold icosahedral symmetry axis. The VP6 densities were averaged by using the 12 matrices and the computer program AVE (17, 18). For the averaging, an ovoid mask with radii of 50 and 60 Å in the two orthogonal directions was defined that generously enveloped the VP6 trimer on the threefold axis. The refined matrix transformations were compared with transformations derived from an atomic model built into the x-ray map (B. McLain, E.S., R.B., and S.C.H., unpublished data). Coordinates transformed by the refined matrices deviated from those transformed by using matrices based on the atomic model with an rmsd of 0.7 Å. This alignment error contributed further to the signal attenuation in the final structure (see below).

The final 13-fold averaged density is shown in Figs. 2e and 6b-d. It has clear features for most of the larger side chains, including aromatic, aliphatic, polar, and charged side chains. The map compares well with an electron density map determined by x-ray crystallography (B. McLain, E.S., R.B., and S.C.H., unpublished data) at 3.8-Å resolution (Fig. 6a), effectively calibrating the resolution of the 13-fold averaged density. The resolution also was estimated by using the Fourier shell correlation (FSC) criterion (Fig. 4), suggesting a resolution of ~4 Å. The interpretation of the FSC curve comparing 13-fold averaged densities of reconstructions using two halves of the data is complicated by the tight masking imposed during the averaging procedure (see Materials and Methods), making the resolution estimate less accurate. The influence of the mask on the FSC was reduced by smoothing the edges of the mask.

**Discussion**

The resolution obtained in the present work represents a substantial step forward in the application of cryo-EM. The promise of achieving a map interpretable by an atomic model has driven the development of new sample preparation techniques, instrumentation, and image processing methods (13, 19–21). An estimate of the number of images that might be required to attain near-atomic resolution was made 12 years ago (22). Subsequent single-particle reconstructions of hepatitis B virus capsids at 7- to 9-Å resolution were adequate to resolve secondary structures and permitted an approximate trace of the subunit polypeptide chain (23, 24). Virus particles are particularly suitable for high-resolution imaging because of their high symmetry, their generally high stability, and their solubility in aqueous buffers. Moreover, their high molecular mass gives rise to strong contrast in the EM, but limitations commonly encountered in cryo-EM applications, such as beam-induced movement and specimen charging, greatly reduce image contrast from the values expected under ideal conditions. Early estimates of the minimum number of images required to obtain a 3-Å resolution structure ranged between 1,400 and 12,600 (22, 25), depending on the assumed

**Fig. 6.** Comparison of x-ray crystallographic and EM density maps. (a) X-ray crystallographic, icosahedrally averaged 2F₀ – F₁ density map at 3.8-Å resolution and atomic model built into this map (B. McLain, E.S., R.B., and S.C.H., unpublished data). The density corresponds to area 2 in Fig. 2c. (b) Same area as in a but showing the cryo-EM map. (c) Cryo-EM density in area 3 (Fig. 2c) showing clearly separated β-sheet strands. (d) Stereo image of area 4 (Fig. 2c) showing amino acid side chains protruding from different α-helices.
signal-to-noise ratio required in an interpretable reconstruction. Both estimates assume physically perfect image formation, without loss of contrast because of the limitations mentioned above. Contrast in experimentally observed images usually is only \( \approx 10\% \) of the contrast expected in a perfect image (22). In our work, the images of 8,400 virus particles contain 8,400 \( \times 60 \times 13 \) = 6.6 million VP6 molecules. Besides the reduced image contrast, inaccuracies in the determination of image defocus and alignment parameters (three angles and \( x, y \) coordinates) also contribute to the requirement for a very large number of particles (26). An estimate of the alignment error (see Materials and Methods) suggests that angles were determined with an error of 0.2° and \( x, y \) coordinates within 0.2 Å. An angular error of 0.2° is expected to attenuate the signal in the reconstruction of a 700-Å particle by a factor of \( \approx 0.3 \), whereas a 0.2-Å \( x, y \) error leaves the signal almost unchanged (26). A second limitation comes from the errors in the measurement of the defocus used to record images. This error has been estimated to be \( \approx 100 \) Å for images containing carbon film and \( \approx 1,000 \) Å for images without carbon. The imaged specimen area included in each micrograph contained 5000 filaments, and the virus particles themselves also contributed significantly to the Thon ring pattern used to determine the defocus. If we assume a defocus error of 300 Å, we expect the signal in the reconstruction to be attenuated by a further factor of 0.4. Other limitations that are not detectable as easily include magnification variations among images (27) and unintentional beam tilt. Finally, the data were not corrected for the curvature of the Ewald sphere, introducing a phase error of \( \approx 60° \) at 3.8-Å resolution (28), which will further attenuate the signal in the reconstruction by a factor of \( \approx 0.5 \). The attenuation therefore is similar to that expected for a defocus error of half the particle diameter (350 Å).

The ability to trace the protein backbone in a structure introduces powerful additional constraints, which have been used in x-ray crystallography to improve the density map and ultimately the coordinates (“refinement”). Now that traceable maps also can be obtained by using single-particle reconstruction from cryo-EM, new refinement methods can be considered. For example, a partial model of a structure could be used to improve the reference used for particle alignment. A partial model also could be used to evaluate the quality of a structure and thus to provide an additional constraint in a refinement or to derive masks that reduce or remove noisy parts of the reconstructed image.

Materials and Methods

EM. Two batches of rotavirus DLPs were prepared as described (29). DLP stock solution was diluted to \( \approx 5 \) mg/ml, and 2 μl of batch no. 1 was applied to lacy carbon grids, whereas 2 μl of batch no. 2 was applied to C-flat holey carbon grids (CF-12-4C; Electron Microscopy Sciences). Images were collected on Kodak ISO163 film by using an FEI F30 electron microscope at 300 kV, \( \approx 59,000 \) magnification, and a Gatan 626 cryo holder. We used an underfocus of 1.1–3.5 μm and an electron dose of 15–20 e/Å. The spot size was set to 6, and a C2 aperture of 50 μm was used, giving a beam cross-over of \( \approx 5 \) nm and divergence of 5 μrad. This results in a spatial coherence width of \( \approx 600 \) Å (30), sufficient to preserve contrast at 3.8-Å resolution and 3.5-μm defocus. The beam pivot points were centered with the specimen at eccentric height, and the beam tilt was aligned by using the objective rotation center. The condenser astigmatism was minimized, followed by coma-free alignment.

These last two steps were repeated several times until there was no further improvement. Wherever possible, the beam was placed symmetrically over a hole in the carbon film to ensure the presence of carbon on all sides of the illuminated area.

A calibrated magnification of 58,080 was used for contrast transfer correction and image processing by FREEALIGN. A more accurate magnification of 56,890 subsequently was determined with a standard (from myosin paracrystal) and used to perform noncoshedral averaging. Comparison with the structure determined by x-ray crystallography gave a magnification of 56,540, close to the calibrated value.

Image Processing. We initially selected 386 micrographs by visual inspection and digitized them using a Zeiss SCI4 scanner with a 7-μm step size, giving a pixel size on the specimen of \( \approx 1.2 \) Å. Processing of the data was accelerated by initial decimation of the images with 2 \( \times \) 2-pixel averaging. Power spectra from each micrograph (including both carbon and ice areas) were calculated to select for micrographs with preserved high-resolution signal. Then, 8,696 and 9,429 particles were selected from images of batch nos. 1 and 2, respectively, by using the computer program SIGNATURE (31). The astigmatic defocus, specimen tilt axis, and tilt angle for each micrograph were determined by using the computer program CTFTILT (32), giving separate defocus values for each particle according to its coordinate in the micrograph. The detected specimen tilt was unintentional and remained \( \approx 5° \) for most micrographs.

By using the search function of the computer program FREEALIGN (13) with a 3° search step, particles collected from batch no. 1 and 2 were aligned against a structure of reovirus (33) filtered at 30-Å resolution and scaled to match the diameter of DLP. This step was followed by 70 cycles of refinement, converging to a structure of DLP. For the refinement, the density corresponding to the RNA genome was masked in the reference structure to enhance the signal of the ordered part of the structure. Particles collected from batch no. 2 were added and aligned to the DLP structure by using FREEALIGN’s search function. Fifteen additional refinement cycles at the full resolution of 1.2 Å/pixel—followed by randomized parameter search (13), another refinement cycle, and a final selection of 8,400 particles with a correlation coefficient \( \approx 0.14 \) resulted in a structure of 5.1-Å resolution before noncoshedral averaging. For the randomized parameter search, a coarse search grid with angular step of 5° was used. This grid search was repeated 50 times for each particle by using random starting points, leading to different searched orientations in each grid search. The resolution was determined by using the FSC criterion and calculated between two reconstructions each containing half of the data and a threshold of 0.143 (34). A negative temperature factor of 450 Å² (x-ray notation) was applied, and the structure was filtered at 3.8-Å resolution. The low-pass filter had a cosine-shaped cut-off with a width of \( \approx 20 \) Fourier pixels.

Estimation of Alignment Errors. A reference structure was calculated with half of the aligned data used to calculate the 5.1-Åicosahedral reconstruction. The reference was not improved with 13-fold noncoshedral averaging. The alignment parameters then were altered by adding small angular perturbations with a uniform distribution of \( \approx 3° \) and small positional perturbations of \( \approx 3.5 \) Å. One cycle of refinement was carried out with FREEALIGN, which brought most of the parameters of particles with a high correlation coefficient (see above) back close to their starting values before perturbation. Some particles were not refined successfully and ended up with a significantly lower correlation coefficient, compared with those refined successfully. These particles, and a small number of other particles with large angular deviations from the starting angles, were not considered further. The standard deviation between the refined and the starting parameters was calculated and used to estimate the parameter errors.

ACKNOWLEDGMENTS. This work was supported by National Institutes of Health Grants GM-62580 (to N.G. and S.C.H.), CA-13202 (S.C.H.), and AI-53174 (to P.R.D. and S.C.H.), an Ellison Medical Foundation New Scholars in Global Infectious Diseases Award (to P.R.D.), and grants from the Health Research Council of New Zealand (to R.B.), N.G. and S.C.H. are Howard Hughes Medical Institute Investigators.