Structural and Electrostatic Characterization of Pariacoto Virus: Implications for Viral Assembly

Batsal Devkota,1 Anton S. Petrov,1 Sébastien Lemieux,2 Mustafa Burak Boz,3 Liang Tang,4 Anette Schneemann,5 John E. Johnson,5 Stephen C. Harvey1,3,6
1 School of Biology, Georgia Institute of Technology, Atlanta, GA 30332-0230
2 Department of Computer Science and Operations Research, Université de Montréal, Montreal, Quebec, Canada H3C 3J7
3 School of Chemistry & Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332-0230
4 Department of Molecular Biosciences, The University of Kansas, Lawrence, KS 66045
5 Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037
6 Parker H. Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA 30332-0230
Received 3 November 2008; revised 19 January 2009; accepted 30 January 2009
Published online 18 February 2009 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bip.21168

ABSTRACT:
We present the first all-atom model for the structure of a T = 3 virus, pariacoto virus (PaV), which is a nonenveloped, icosahedral RNA virus and a member of the Nodaviridae family. The model is an extension of the crystal structure, which reveals about 88% of the protein structure but only about 35% of the RNA structure. New modeling methods, combining coarse-grained and all-atom approaches, were required for developing the model. Evaluation of alternative models confirms our earlier observation that the polycationic N- and C-terminal tails of the capsid proteins must penetrate deeply into the core of the virus, where they stabilize the structure by neutralizing a substantial fraction of the RNA charge.
This leads us to propose a model for the assembly of small icosahedral RNA viruses: nonspecific binding of the protein tails to the RNA leads to a collapse of the complex, in a fashion reminiscent of DNA condensation. The globular protein domains are excluded from the condensed phase but are tethered to it, so they accumulate in a shell around the condensed phase, where their concentration is high enough to trigger oligomerization and formation of the mature virus.
Keywords: viral assembly; RNA virus; coarse-grain; modeling

INTRODUCTION
Pariacoto virus (PaV), a T = 3, nonenveloped, icosahedral virus, is a member of the Nodaviridae family. It was originally isolated in Peru from the Southern Armyworm, Spodoptera eridania.1 Its genome consists of two positive-sense ssRNAs.2 RNA1 (3011 nucleotides) codes for protein A, the catalytic subunit for the RNA-dependent RNA replicase. RNA2 (1311 nucleotides) codes for capsid precursor protein α. One hundred and
eighty of these $\alpha$ proteins and the genome assemble together to make up the virus. Upon maturation, the 401-residue protein $\alpha$ is cleaved between Asn 361 and Ser 362 to give proteins $\beta$ and $\gamma$. Ever since it was isolated, PaV has been extensively studied using various techniques.\(^3\)\(^-\)\(^6\) The relatively small size (diameter 30 nm) and the ease with which it can be produced in various cell lines\(^7\) make PaV and other members of the Nodaviridae family relatively easy to characterize at the molecular level.\(^8\)\(^-\)\(^10\)

Structural studies of viruses are very important to understand protein-protein and protein-RNA interactions as well as to understand assembly pathways in RNA viruses.\(^11\)\(^-\)\(^14\) In the last few years, many studies have been done on RNA viruses using molecular modeling as a supplementary method when other methods such as x-ray crystallography and cryo-electron microscopy (cryo-EM) do not give sufficient structural information. An all-atom model was derived for satellite tobacco mosaic virus, a $T = 1$ virus, using molecular modeling.\(^15\) Those authors also carried out molecular dynamics (MD) simulations on the model to study the stability of the protein capsid and the RNA genome.\(^15\) Electrostatic interactions between RNA and the protein capsid were studied in cowpea chlorotic mottle virus by modeling the virus using coarse-grained modeling and representing RNA nucleotides by unconnected spheres that were distributed using the Monte Carlo method.\(^16\) In that study, no attempt was made to model the RNA structure. Other electrostatic studies of RNA viruses have also been aimed at understanding the molecular interactions and their effects on virus structure.\(^17\)\(^,\)\(^18\)

The 3.0 Å x-ray crystal structure of PaV revealed an asymmetric unit with three quasi-equivalent protein subunits (A, B, and C) and one strand of a 25 base pair RNA duplex.\(^6\) Residues 7–378 and 394–401 are seen in subunit A; 49–382 in subunit B; and 51–382 in subunit C. In all three subunits, there is a clear break at the postassembly cleavage site between residues 361 and 362, consistent with the detection of protein $\gamma$ in purified PaV proteins.\(^3\) Sixty asymmetric subunits combine to form the icosahedral capsid, with 30 RNA duplexes lying along subunit contacts across the icosahedral twofold axes, forming a dodecahedral cage inside the capsid. Residues 83–321 of the A, B, and C subunits are folded into an eight-stranded antiparallel $\beta$-sandwich, similar to proteins in other nodaviruses. Complementing the x-ray studies, cryo-EM showed the overall structure of PaV at 23 Å resolution, which matched well with a low resolution model derived from the atomic coordinates.\(^6\) Cryo-EM also confirmed that the part of the RNA genome that was resolved in the x-ray structure forms the edges of a dodecahedral cage inside the capsid.

Although x-ray crystallography and cryo-EM have provided significant information regarding the structure of PaV (6), the complete atomic structure has not been determined. RNA at the dodecahedral edges accounts for only 35% of the total genome. The remaining 65% of the RNA lies inside the dodecahedral cage and was not resolved in the crystal structure because it lacks icosahedral symmetry. In addition, the twenty vertices at which the RNA duplexes are connected could not be resolved, presumably because different vertices have different structures. Similarly, protein subunit A is missing six residues at the N terminal end and 15 at the C terminal in the crystal structure, while the B and C subunits are both missing about 50 residues at the N-terminus and 19 residues at the C-terminus in the crystal structure.\(^6\)

In this paper, we report a model for the complete virus, examining the interactions of the basic N-terminal tails with the RNA genome, and their role in the stability of the mature virus. We built a model of the missing 65% of the genome and the unresolved protein residues with coarse-grained modeling. After refinement of the coarse-grained model, it was reconverted to an all-atom representation using special algorithms. We generated two all-atom models for the virus that differed in the conformations of the N-terminal protein tails and the extent to which they penetrate into the RNA genome. We tested these against the experimental radial density distributions from cryo-EM, and we evaluated the relative stabilities of the two models by comparing their energies. The result is the first all-atom model for a complete $T = 3$ virus. Further, this effort has led to a new model for the assembly of small, nonenveloped icosahedral RNA viruses.

**RESULTS**

Initial modeling was done using a coarse-grained approach, in which each residue is represented by one bead (pseudo-atom). These are located at the positions of the phosphate groups for the RNA, and at the alpha carbons for the protein. When the complete virus had been modeled this way, the model was converted to an all-atom representation for final refinement.

The modeling of the Pariacoto virus genome posed several challenges because of the limited amount of available structural data. To begin with, the secondary structure for the PaV genome is not known. We used a hypothetical secondary structure mapped onto the dodecahedral cage (see Figure 1). This is the same secondary structure that we proposed earlier.\(^19\) Those parts of the RNA genome that do not form the edges of the dodecahedral cage occupy the central cavity of the virus and are connected to the cage by RNA “stalactites” that drop from the cage towards the center of the virus. The
The exact number of stalactites is not known, but the secondary structure model (see Figure 1) dictates a combination of three-way junctions and four-way junctions as structural motifs connecting the RNA on the dodecahedral cage with the RNA in the interior (see Figure 2). Nothing is known about the RNA structure in the interior, so we have to postulate plausible structures for the stalactites. We used twelve copies of a structure derived from domain IV (residues 1764–1988) of the *E. coli* 23S RNA to represent these. There is nothing significant about this choice except that it provides a fragment of 225 residues of known structure. All twelve stalactites had identical starting structures, but they have significantly different conformations in the final model (model_8), as a consequence of energy minimization to eliminate steric conflicts (Supporting Information Figure 1).

We emphasize that this is a first-generation model for the PaV genome. Because of the limited experimental data, we have focused on developing a plausible model rather than a rigorous one. The arrangement of the secondary structure (see Figure 1) is only one of many possible ways to arrange the RNA to generate the required 30 double helices on the edges of the dodecahedral cage. The number and locations of the stalactites are not known, nor is there any experimental information on the RNA in the interior, beyond the radial density distribution. As a consequence, this should be considered a representative model from a very large ensemble of possibilities.

The proteins in the model come directly from the crystal structure, with the exception of those residues that are not seen in the crystal structure (1–6 and 379–393 in subunit A; 1–48 and 383–401 in subunit B; and 1–50 and 383–401 in subunit C). We initially modeled the missing residues using a coarse-grained approach RNA (Supporting Information Figure 2). The N- and C-terminal tails contain numerous basic residues, which serve to partially neutralize the charge on the genomic RNA, so our modeling protocol was aimed at finding conformations for these tails that would maximize their interactions with the RNA. The final conformations of the

---

**FIGURE 1** Secondary structure map for the Pariacoto virus genome, adapted from our earlier model. This is just one solution for mapping the RNA onto the dodecahedral cage in a fashion that is consistent with the data from x-ray crystallography and cryo-EM. The RNA secondary structure is not known, nor is it known whether all base pairing in RNA1 and RNA2 is intramolecular, or whether there are base pairs between RNA1 and RNA2. In the absence of such information, we have represented the bipartite genome as one single strand. The 5’ and 3’ ends are indicated. Each edge of the dodecahedral cage contains an antiparallel RNA double helix. Red and yellow circles are three- and four-way junctions, respectively. These junctions identify those vertices where RNA "stalactites" drop down into the interior of the virus to connect with that part of the RNA not on the dodecahedral cage.

**FIGURE 2** Stereo images of model RNA junctions at the vertices. (a) Three-way junction at a vertex with no stalactite. The three RNA duplexes lie on three edges of the dodecahedral cage. (Members of this class of three-way junction occur on unlabeled vertices in Figure 1.) (b) Three-way junction with two duplexes on edges of the dodecahedral cage and the third duplex (green and yellow) forming a stalactite that drops into the interior of the virus. The third edge of the dodecahedral cage is occupied by a stem-loop (red) that comes from a neighboring vertex. (Members of this class of three-way junction are located at vertices indicated by red dots in Figure 1.) (c) Four-way junction connecting three RNA duplexes on the edges of the dodecahedral cage, and a stalactite. (Four-way junctions are located at vertices indicated by yellow dots in Figure 1.)
protein tails vary significantly in the final model (model_8). A section through the final model is shown in Figure 3.

We examined the role of the amino- and carboxy-terminal protein tails in stabilizing the virus by developing two different models (Materials and Methods). In one (model_12), the protein tails are largely confined to a region just under the capsid, while in another (model_8), they penetrate deeply into the interior of the virus. These differences are reflected in different radial density distributions (see Figure 4). In model_8, the protein tails contribute significantly to the broad density peak at a radius of about 50 Å (Figure 4a). Peaks around this radius have been found in PaV (Figure 4b) and in other nodaviruses. This is also consistent with density maps in flock house virus (FHV), which is closely related to PaV. The radial density distribution for wild-type FHV has a broad peak at $R \sim 32$ Å, but that peak is missing in mutant FHV in which 30 amino acids have been deleted from the amino terminus. 19

Electrostatic calculations on the two models show that model_8 is energetically much more favorable than model_12. The electrostatic interaction energy between the RNA and the PaV capsid is much lower for model_8 (−3910 kcal/mol) than for model_12 (−523 kcal/mol). This is easily explained: protein tails that penetrate deep into the core of the virus stabilize PaV by neutralizing a large fraction of the charge of RNA genome.

Figure 5 depicts the electrostatic potential mapped onto the solvent-accessible surface area of PaV. The external surface of PaV is almost neutral (Figure 5a), whereas the interior of the virus bears both positive charges (the protein tails) and negative charges (RNA). The lower panels of Figure 5 show the potential calculated for the virus without (Figure 5c) and with (Figure 5d) RNA, mapped onto the surface of the empty capsid. The positively charged tails (blue in Figure 5d) are fully neutralized and even reveal some negative potential on their surface due to the close proximity of RNA (Figure 5d). This is because the total charge of RNA is almost a factor of two greater than that of the capsid.

FIGURE 3  A 20 Å section through the center of the final model (model_8). The thumbnail sketch at upper left shows the location and thickness of the slice by omitting the capsid proteins (dark gray) in the slice. In the main figure, those parts of the capsid proteins whose structures were restrained to the crystallographic conformation are shown in blue; they have icosahedral symmetry. Our model for those parts of the protein tails whose structures were not resolvable in the crystal structure are shown in red; these do not have icosahedral symmetry. RNA is shown in green.

DISCUSSION

There are three pieces of evidence that the polycationic protein tails penetrate deeply into the interior of nodavirus genomes. First, mutant FHV that lack 30 N-terminal amino acids are missing the broad 32 Å peak seen in the cryo-EM radial density distribution for wild-type FHV. 19 Second, electrostatic calculations show that deep penetration of the tails has a stabilizing effect, because of greater neutralization of the RNA charge. Finally, the radial density distribution for...
our model_8 is somewhat closer to the experimental radial density distribution than that of model_12, and tails in the former penetrate much deeper into the genome than those in the latter model.

This observation has important implications for viral assembly.

The assembly of small icosahedral RNA viruses like PaV and FHV is quite different from bacteriophage assembly. Interactions between phage capsid proteins are strong enough that capsids assemble spontaneously. The DNA genome is then forced into the preformed capsid by an ATP-dependent motor; there is little or no attraction between the DNA and the capsid proteins. This promotes ejection of the genome upon infection of the host bacterium. In contrast, protein-protein interactions are weak in nodaviruses, and RNA-protein interactions are strongly attractive. As a consequence, capsids are only formed when the proteins and RNA are both present, and empty capsids do not exist.

We propose a simple mechanism for the assembly of nodaviruses. Positively charged protein tails bind to the RNA (Figure 6a), with RNA replication, protein synthesis and RNA-protein binding occurring very closely in time and space. When a sufficient quantity of the RNA charge is neutralized, the resulting complex collapses in a process reminiscent of DNA condensation (Figure 6b). We believe that most of these interactions are nonspecific, although in the mature virus there is evidence of a specific interaction between RNA2 and the N-terminal tail. In addition, the crystal structure shows ordered interactions between the RNA and 36 N-terminal residues of subunit A, and between the RNA and eight residues of the C-terminus of subunit A, although the RNA sequence involved in these interactions cannot be determined. We hypothesize that the globular domains of the capsid proteins are squeezed to the outside of the collapsed state, as shown in Figure 6b. This provides a sufficiently high local concentration that the relatively weak protein-protein affinity is sufficient to cause oligomerization, leading to formation of the mature capsid (Figure 6c). In FHV-infected cells, all these interactions appear to occur in a confined environment created by alterations induced in the mitochondrion by the viral polymerase.

One remarkable observation suggests that this mechanism might apply to many single-stranded viruses. Belyi and
Muthukumar examined 16 wild-type and three mutant viruses (both DNA and RNA viruses) with genomes ranging from about 1 to 12 kb. They found that the ratio of the genome size to the net charge on the terminal protein tails is 1.61 ± 0.03, an unexpectedly uniform ratio. Such a narrow range might be explained by our model, because the initial collapse would require sufficient charge neutralization to overcome RNA-RNA repulsions, but not so much as to lock the condensed state into a rigid, fixed configuration that might inhibit assembly of the mature capsid structure around the condensed mass.

This model provides a simple mechanistic basis for explaining how the relatively weakly associating proteins can force RNA into a small compact volume: the very strong electrostatic interactions between the polyanionic RNA and the polycationic protein tails provide a sufficiently favorable change in enthalpy to overcome the unfavorable entropic penalty associated with the dramatic reduction in RNA conformational space.

In summary, we present the first all-atom model of a complete T = 3 virus. New approaches, combining coarse-grained and all-atom modeling, were developed during this work. These should be useful in modeling other RNA and DNA viruses. The model is consistent with all the available experimental data, and it is sterically plausible, although it is by no means a unique model. Most important, it leads to a simple mechanistic hypothesis for the assembly mechanism of small nonenveloped RNA viruses. It will be exciting to test this model both experimentally and using computer simulations.

**MATERIALS AND METHODS**

**General Approach**

The model was built in four stages. First, pieces of the RNA model were developed in all-atom detail. The 750 nucleotides of the dodecahedral cage came directly from the crystal structure. The three-and-four-way junctions at the vertices were developed independently in all-atom detail, then connected to the cage. Second, the full RNA model was developed by joining these pieces together and attaching the stalactites containing the remainder of the RNA, as described in the next paragraph. A coarse-grained approach was used throughout this stage: each nucleotide is represented by a single large bead (pseudoatom), located at the position of the phosphate group. A more complete description of this "all-P" representation is available elsewhere, along with a full description of the corresponding force field. Third, the protein model was developed, again using a coarse-grained approach, with one pseudoatom for each amino acid. Those residues that are revealed in the crystal structure were held fixed, while the amino and carboxy-terminal tails of the proteins were pulled toward the center of the capsid to neutralize the RNA. Refinement of the coarse-grained models used a combination of MD and energy minimization, to idealize the geometry and eliminate serious steric conflicts. The fourth stage of modeling consisted of conversion of the coarse-grained model to an all-atom representation, followed by gentle optimization using a combination of MD and energy minimization.

The volume inside the dodecahedral cage is too small to attach stalactites coming inwards from each of the twelve vertices without interpenetration. To solve this problem, we expanded the diameter of the dodecahedral cage by a factor of two, which increased the overall volume of the dodecahedral cage eightfold. This allowed us to add the stalactites from the vertices without interpenetration. The expanded model was then contracted to the actual size in twelve scaling steps, with extensive energy minimization at each step (see Figure 7). A similar expansion-contraction cycle is required to thread the protein tails into the crowded interior of the virus (Supporting Information Figure 2).

**RNA Modeling**

**Coarse-Grained RNA Modeling.** The crystal structure for the asymmetric unit of PaV (1F8V.pdb) was taken from the RCSB Protein Data Bank. The dodecahedral RNA cage was generated by applying the BIOMT TRANSFORMATION matrix given in the file, using the oligomer generator tool in the VIPER database. The vertex structures were defined by the secondary structure (see Figure 1). Each vertex has either three or four extensions of RNA coming out of it (see Figure 2). Short stems were added at twelve vertices, as stubs to which the stalactites were subsequently added. Twenty non-RNA "vertex pseudoatoms" were added at the vertices of the dodecahedral cage, to form a framework that could be easily expanded and contracted.

The size of the framework was increased by doubling the values of the x, y, and z coordinates of each vertex pseudoatom. We cut the RNA duplex on each edge in half, scaling coordinates to move each half-duplex to the appropriate vertex of the expanded framework, along with the connecting RNA vertex structures and the associated stubs. Stalactites were then attached to each of the twelve stubs. This initial model contained all 4322 RNA residues, plus twenty cage pseudoatoms to define the vertices (Figure 7a).

The edges of the dodecahedral cage were decreased to the original length in multiple steps, decreasing the ideal bond length \( b_0 \) of the framework by 5 Å in each step and minimizing until convergence (see Figure 7). The minimization was done using our in-house molecular mechanics package, YAMMP. The energy terms used in the minimization are given in Supporting Information Table 1. Since all the terms used in the potential energy function of all-P models are harmonic, minimization of the model should lead to zero energy, if all restraints can be satisfied.

During minimization, the stalactites RNA were relatively free to move and adjust their conformations, to avoid steric overlap. This was achieved by giving the intrastalactite energy terms softer force constants than those for the RNA duplexes on the dodecahedral cage (Supporting Information Table 1). The latter were restrained by using strong force constants in the energy terms, and by the addition of pseudobonds, angles and improper torsions connecting each vertex pseudoatom to the ends of the RNA duplexes at each edge. Duplexes on the dodecahedral cage did not deviate significantly from the crystal structure during the contraction/minimization cycles, and they were returned to the crystallographic conformation on conversion of the coarse-grained model back to all-atom form.
Conversion of Refined Coarse-Grained RNA Model to All-Atom Model

Generating an all-atom model from phosphate positions is a challenging problem. The bond and angle restraints in the all-P models are based on observed distributions of P-P distances and P-P-P angles in the Nucleic Acid Database. With only these restraints, groups of four or more successive P atoms in any all-P model may have conformations that differ somewhat from those in real RNA structures. As a consequence, all-atom models can be generated fairly easily in double-helical regions, but all-atom models for other regions (loops, bulges, single-strands) are necessarily more speculative. This is not inappropriate, considering the modesty of our overall goal: generate a plausible RNA model, in terms of the connections along the backbone and the absence of serious steric problems. Here we discuss how plausible all-atom models can be built for all regions of the all-P model.

Briefly, the procedure used here builds all-atom models using a database of nucleotide conformations derived from all RNA-containing structures in the PDB as of April, 2006. In base-paired regions, four phosphate positions (0 and +1 on each strand) serve as anchor points, and a pair of nucleotides from the database must be fit to the structure, one on each strand. In nonbase-paired regions, the four anchor phosphates are those −1, 0, +1, and +2 relative to the nucleotide being placed. The compatibility of all examples in the database with a particular position is assessed by requiring that the base be identical to the one being modeled, and that the root mean square deviations of the four phosphate positions in the example be within 1.5 Å of the anchor phosphates in the all-P model. Only examples that pass this compatibility test are kept within the search space of each nucleotide.

The modeling problem then becomes one of exploring the search space of the whole molecule to determine which combination of examples gives the most plausible structure, where plausibility is defined as the lowest energy (van der Waals plus electrostatics, using the AMBER 8 force field). This optimizes base pairing and stacking, while minimizing steric clashes. Searching is done in a piecewise fashion, focusing on individual regions, to optimize performance. The most plausible structure is then refined by optimization of the ribose conformations, followed by energy minimization and a short annealing of the entire model, using MD.
**Protein Modeling**

For modeling the missing protein residues, we followed a similar methodology as in the case of RNA modeling, expanding the capsid, adding missing amino acids, and then shrinking the capsid back to its original size in multiple steps, with minimization at each step.

**Coarse-Grained Protein Modeling**

First, the capsid was expanded three times in length by simply multiplying the coordinates for the capsid proteins by 3. The crystallographic residues facing towards the RNA were converted into a model where two consecutive residues were represented by a single pseudoatom (2C-model). The rest of the crystallographically-resolved protein domains were represented by twelve pseudoatoms each, defining the face, edge and vertices of the equilateral triangle of each asymmetric unit. The missing N-terminal residues were generated in extended linear form pointing towards the RNA genome at the center. C-terminal residues were generated as a random coil. Residues for both the N- and C-terminal tails were represented by one pseudoatom per residue (Supporting Information Figure 2).

The starting capsid model was scaled back down to the original size in a series of steps, testing different scaling ratios and van der Waals (vdw) diameters for the pseudoatoms of the protein tails. We examined scaling ratios between 0.95 and 0.99, finding that different scaling ratios did not significantly affect the configurations of the protein tails (data not shown). However, changing the vdw diameters from 8 to 12 Å significantly affected the penetration of the protein tails into the RNA genome, as seen in the radial density distribution (Figure 4a). The resulting structures, designated model_8 and model_12, have dramatically different conformations for the protein tails. In model_8, the tails penetrate deeply into the RNA core, while they lie mostly on the outside of the RNA core in model_12.

**Conversion of Refined Coarse-Grained Protein Model to All-Atom Model**

Model_8 and model_12 were converted into all-atom representation using PULCHRA. PULCHRA converts Cx models to all-atom models using a rotamer library prepared from the statistics of Cx distances in the Protein Data Bank. The complete all-atom models, including all residues of the RNA genome and the capsid proteins, were energy minimized with NAMD, using the CHARMM force field.

**Electrostatic Calculations**

Calculations of the electrostatic potential were performed using the Adaptive Poisson-Boltzmann Solver. CHARMM27 force field radii and charges were assigned to the minimized all-atom structures of Model_8 and Model_12 using the PDBPQR routine, yielding a charge of \(+46e\) for each of the 60 capsomers and \(-4320e\) for the RNA genome, where \(e\) is the charge on the proton. This resulted in a net charge of \(-1560e\) for the complete virus. The nonlinear version of the Poisson-Boltzmann equation was solved numerically on the 225 × 225 × 225 grid with an initial grid spacing of 2.0 Å, followed by focusing with the grid spacing reduced to 1.5 Å. The dielectric constants of the interior and exterior of the macromolecules were set to 10 and 78.5, respectively. The ionic strength was set to 100 mM, using only monovalent ions. The resulting potentials were visualized using Chimera by mapping them onto the solvent-accessible surface of the models generated at the coarse-grained level.

**Model Evaluation**

The coarse-grained pseudoatomic model of the genome was checked for the presence of possible knots using the “knot” program. Our RNA model does not contain any knots. The all-atom genome model reconstructed from the pseudoatomic model was checked for interpenetration of rings and correct stereochemistry using PROCHECK, provided in the RCSB PDB website (http://www.pdb.org). There are no ring penetrations or other stereochemical problems. The RNA and protein distributions inside the complete all-atom models of the virus were compared with the native virus by generating density maps and corresponding radial density distribution functions (see Figure 4) from the final all-atom models using SPIDER.

SCH is a Georgia Research Alliance Eminent Scholar and is partially supported by the Parker H. Petit Institute of Bioengineering and Biosciences at the Georgia Institute of Technology. The authors thank Vijay Reddy, Robert K.-Z. Tan and Thomas R. Caulfield for discussions and valuable insights.

**REFERENCES**


Reviewing Editor: Sarah Woodson