Screening for the Location of RNA using the Chloride Ion Distribution in Simulations of Virus Capsids

Daniel S. D. Larsson and David van der Spoel*

Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, Box 596, SE-751 24 Uppsala, Sweden

ABSTRACT: The complete structure of the genomic material inside a virus capsid remains elusive, although a limited amount of symmetric nucleic acid can be resolved in the crystal structure of 17 icosahedral viruses. The negatively charged sugar—phosphate backbone of RNA and DNA as well as the large positive charge of the interior surface of the virus capsids suggest that electrostatic complementarity is an important factor in the packaging of the genomes in these viruses. To test how much packing information is encoded by the electrostatic and steric envelope of the capsid interior, we performed extensive all-atom molecular dynamics (MD) simulations of virus capsids with explicit water molecules and solvent ions. The model systems were two small plant viruses in which significant amounts of RNA has been observed by X-ray crystallography: satellite tobacco mosaic virus (STMV, 62% RNA visible) and satellite tobacco necrosis virus (STNV, 34% RNA visible). Simulations of half-capsids of these viruses with no RNA present revealed that the binding sites of RNA correlated well with regions populated by chloride ions, suggesting that it is possible to screen for the binding sites of nucleic acids by determining the equilibrium distribution of negative ions. By including the crystallographically resolved RNA in addition to ions, we predicted the localization of the unresolved RNA in the viruses. Both viruses showed a hot-spot for RNA binding at the 5-fold symmetry axis. The MD simulations were compared to predictions of the chloride density based on nonlinear Poisson—Boltzmann equation (PBE) calculations with mobile ions. Although the predictions are superficially similar, the PBE calculations overestimate the ion concentration close to the capsid surface and underestimate it far away, mainly because protein dynamics is not taken into account. Density maps from chloride screening can be used to aid in building atomic models of packaged virus genomes. Knowledge of the principles of genome packaging might be exploited for both antiviral therapy and technological applications.

INTRODUCTION

Many structures of icosahedral viruses only model the coat protein (CP), and very little is known about the atomic structure of the genome. The crystallographic methods used in the experiments rely on the packing of millions of virus particles in crystals where each individual virion can be positioned in 60 different orientations. Even if the genome has a well-defined structure, X-ray scattering data will only reveal densities in areas where a large fraction of the different orientations agree, and structures with a low degree of icosahedral symmetry are blurred by averaging. It has been suggested that single-particle imaging techniques could overcome this problem;[1] however, the question of what resolution eventually can be obtained by such techniques is far from resolved.[2,3]

Seventeen structures of icosahedral viruses in the PDB contain modeled nucleic acids (summarized in Table 1). For most of them weak densities necessitated the use of arbitrary nucleotide sequences. Common for all these viruses is that they are naked (not membrane enveloped) and have single stranded DNA (ssDNA) or RNA (ssRNA) genomes. The nucleic acids in these structures are always in close association with the interior face of the capsids, which are usually lined with basic side chains. Several RNA viruses also have a flexible and positively charged N-terminal arm that extends into the capsid interior and presumably binds the RNA.[24,25] The two closely related ssDNA bacteriophages α3 (see Table 1 for abbreviations and references) and ϕX174 instead encode for a positively charged peptide, known as the J peptide, that binds to the DNA. Similarly, the three RNA viruses FHV, PaV and PrV also

have positively charged γ peptides binding to the RNA, which are cleaved off of the coat protein.

In most of the structures, the resolved nucleic acids form well-defined secondary structure motifs. A common motif is a double helix on the 2-fold icosahedral symmetry axis parallel to the capsid surface (STMV, PaV, FHV, BBV, and PrV). In PaV, the helix extends all the way to the 5-fold axes and forms a complete dodecahedron cage that accounts for a third of the genome. Another common motif consists of stacked nucleotides forming a single-stranded helix (CCMV, TAV, PhMV, TYMV). In BPMV an extension of the helix reaches across to the symmetry neighbor and forms a trefoil pattern around the 3-fold axis. The CPs of MS2 form dimers in solution, which bind a specific hairpin structure from the genome; such a dimer therefore acts as a self-regulatory inhibitor of CP expression.[9] Crystals of MS2 capsids supplemented with this operator sequence bind the hairpin structure at a majority of the 90 binding sites and they show up clearly in the diffraction structure. The visible DNA in the two closely related viruses MVM and CPV share a conserved fold, a compact structure with stacked nucleotides that folds back onto itself. The DNA in α3 and ϕX174 only diffracts weakly and the model mainly consist of phosphate and sugar backbone fragments of unconnected residues.

Simulations of coarse-grained ribonucleic acids modeled as partially charged spheres moving in a static capsid cage have

Table 1. Icosahedral Virus Structures in the Protein Data Bank Containing DNA or RNA

name	PDB ID	resolution (Å)	type	group	N^a (nt)	S^b	total (Kb)	%
bean pod mottle virus (BPMV)[4,5]	1BMV	3.00	RNA	$T = 3$	33	20	3.7^c	18
black beetle virus (BBV)[6]	2BBV	2.80	RNA	$T = 3$	20	30	4.5	13
cowpea chlorotic mottle virus (CCMV)[7]	1CWP	3.20	RNA	$T = 3$	10	60	3.0^d	20
Desmodium yellow mottle virus (DYMV)[8]	1DDL	2.70	RNA	$T = 3$	9	60	6.3	9^e
enterobacteria phage MS2 (MS2)[9]	1ZDH	2.70	RNA	$T = 3$	13	90	3.6	33^f
Flock house virus (FHV)[10]	2Z2Q	2.70	RNA	$T = 3$	24	30	4.5	16
Pariacato virus (PaV)[11]	1F8V	3.00	RNA	$T = 3$	50	30	5.8	35
Physalis mottle virus (PhMV)	2XPJ	3.40	RNA	$T = 3$	3	60	6.7	3
Providence virus (PrV)[12]	2QQP	3.80	RNA	$T = 4$	8	30	8.7	3
satellite tobacco mosaic virus (STMV)[13]	1A34	1.81	RNA	$T = 1$	20	30	1.1	62^f
					+1	60		
satellite tobacco necrosis virus (STNV)[14]	3S4G	6.00	RNA	$T = 1$	7	60	1.2	34^{e,f}
tomato aspermy virus (TAV)[15]	1LAJ	3.40	RNA	$T = 3$	3	60	3.2^d	6
turnip yellow mosaic virus (TYMV)[16]	2FZ2	2.90	RNA	$T = 3$	3	60	6.3	3
canine parvovirus (CPV)[17,18]	1IJS	3.25	DNA	$T = 1$	11	60	5.3	12
enterobacteria phage alpha3 (α3)[19]	1M06	3.50	DNA	$T = 1$	10	60	6.1	10^e
enterobacteria phage phiX174 (ϕX174)[20,21]	1RB8	3.50	DNA	$T = 1$	10	60	5.4	11^e
minute virus of mice (MVM)[22,23]	1Z1C	3.50	DNA	$T = 1$	22	60	5.1	26

^{a}Number of modeled nucleotides. ^{b}Symmetry induced multiplicity of the nucleic acid fragment. ^{c}Bipartite genome; crystal only had the shorter strand. ^{d}Average per virion for multipartite genomes. ^{e}Authors state that nucleotides were built on weak densities and may have partial occupancy. ^{f}Nongenomic RNA.

been used to suggest that the steric envelope and the electrostatic properties of the capsid shell encode much of the information of how the nucleic acid is packaged.[26,27] Here we adapted this concept into a novel approach to elucidate the structure of the genome organization in viruses by studying the ionic atmosphere in and around two virus capsids using explicit negative and positive ions while letting the RNA binding N-termini of the CPs coequilibrate with the ions. We have simulated half capsids of STNV and STMV in atomistic detail, both with and without the known parts of the RNA genome. Four simulations were performed, each of 200 ns length with close to 1 million atoms (Table 2). A half capsid is sufficient to

Table 2. Composition of Simulation Systems

	STMV		STNV	
	without RNA	with RNA	without RNA	with RNA
proteins molecules	30	30	30	30
RNA molecules	0	40	0	60
calcium ions	0	0	61	61
sodium ions	827	887	852	852
chloride ions	1127	827	1214	1064
water molecules	279 183	275 415	283 716	281 376
no. of atoms (total)	913 483	914 499	944 535	943 875

represent all local constellations of protein and RNA molecules and promotes fast equilibration of the ions. We found a clear correlation between the distribution of chloride ions and the position of the resolved nucleotides. The MD simulations were compared to electrostatic calculations by solving the nonlinear Poisson−Boltzmann equation (PBE).

The two viruses STMV and STNV represent excellent model systems for simulations to better understand the packaging of viral genomes. They are both relatively small and represent two different nucleic acid packing motifs. STMV is also the virus where the largest fraction of the packaged RNA is visible (∼62%), and for STNV, low-resolution neutron scattering data
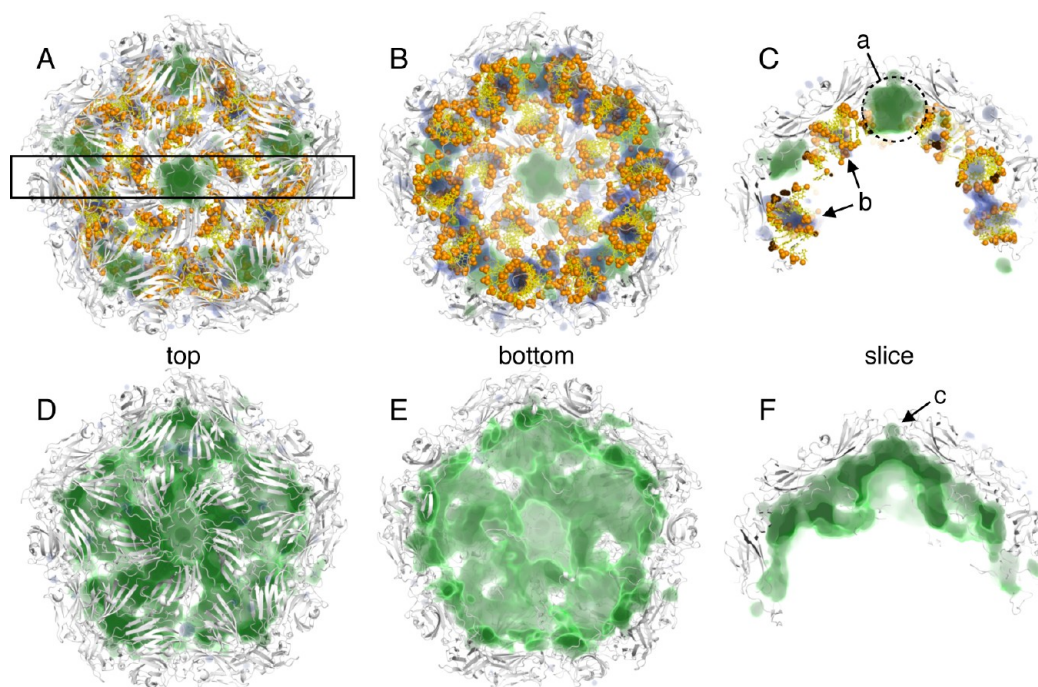
is available,[28] complementing the X-ray data used to prepare the starting structures.

■ RESULTS

The chloride and sodium ions from the final 100 ns of the simulations of the capsid half-shells of STMV and STNV were converted into time-averaged density maps (see the Methods section for details). The extended sampling in the long simulation trajectories allowed analysis of the maps without the need to average over the central 5-fold symmetry. The maps were visualized in PyMOL by coloring regions with high average ion densities. Figures 1 and 2 display the difference maps overlaid onto the protein and, in relevant figures, the RNA structures from the final frames of the simulations. The densities displayed a high degree of icosahedral symmetry showing that the capsids induced a strong icosahedral symmetry on the ion distributions despite the flexibility of the protein side chains and N-termini. The symmetry of the densities gradually falls off toward the rim of the half-shells, as expected.

The densities from the simulations containing the RNA fragments were essentially subsets of the densities in the simulations of the empty capsid half-shells (Figures 3 and 4), highlighting that the chloride density in regions where no RNA was modeled in the crystal structures is indicative of RNA encapsidation potential.

The equilibration of the chloride ions was verified by calculating the autocorrelation function (ACF) for chloride ions bound to lysine and arginine side chains for the empty capsids (Figure 5). The integral of the ACF was used to estimate the correlation time for bound chloride ions as 0.22 ns in STMV and 0.16 ns in STNV. Since the length of the production simulations was 100 ns, we conclude that exchange of ions is fast on the time scale of the simulation.

Satellite Tobacco Mosaic Virus. In the STMV structure, the majority of the visible nucleotides form 9 base-pair-long double helices (A-form like) below the inner capsid surface on the icosahedral 2-fold symmetry axis. Each strand extends one

Figure 1. MD simulation of STMV half-capsid with RNA (top row, A−C) and without RNA (bottom row, D−F). Proteins are shown as white cartoons and RNA as yellow sticks with phosphate groups as orange spheres. Regions with high ion concentration (>4 × 10⁻⁴ Å⁻³) are shown in green (chloride) and blue (sodium). Views from above (A, D), below (B, E), and a slice through (C, F) the half-capsid. The box in panel A indicates the view in panels C and F. Circles and arrows are explained in the text.
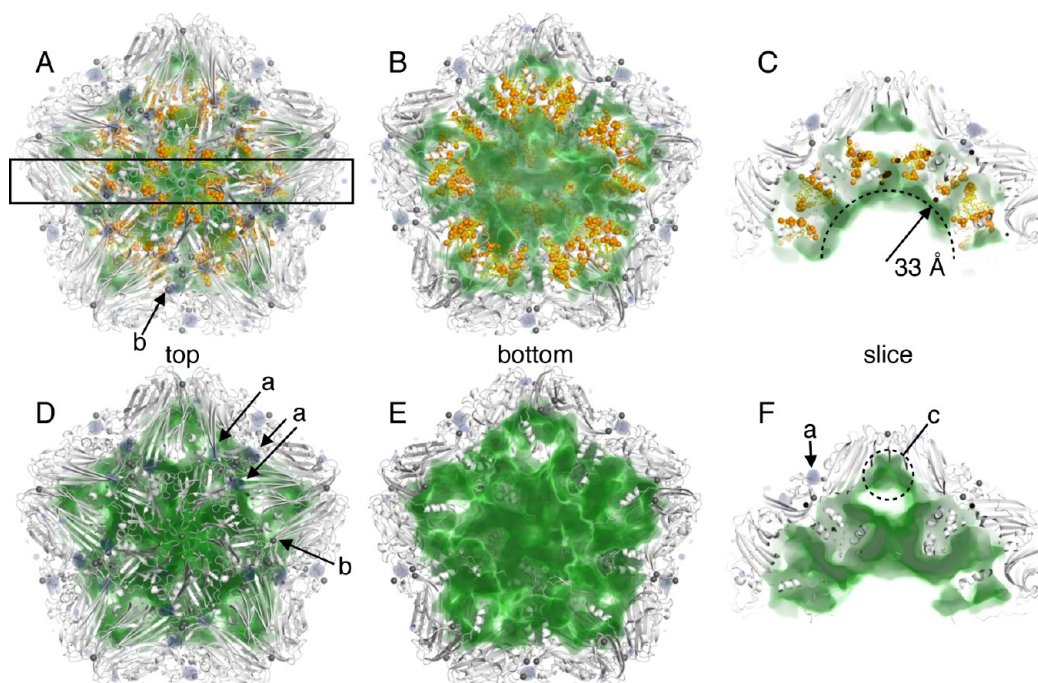

Figure 2. MD simulation of STNV half-capsid with RNA (top row, A−C) and without RNA (bottom row, D−F). Proteins are shown as white cartoons, RNA as yellow sticks with phosphate groups as orange spheres, and calcium ions as gray spheres. Regions with high ion concentration (>4 × 10⁻⁴ Å⁻³) are shown in green (chloride) and blue (sodium). Views from above (A, D), below (B, E), and a slice through (C, F) the half-capsid. The box in panel A indicates the view in panels C and F. The circles and arrows are explained in the text.

additional base in the 3' end. In addition, the phosphate and ribose of an additional nucleotide can be found in a general position. (This fragment was not included in the simulations, however, since the base was missing.)

The chloride ion distribution in the simulation of the RNA-free capsid showed a remarkably high degree of overlap with the crystallographic RNA when superpositioned. The double helix on the 2-fold axis fits very well inside the chloride ion density that radiates out like twisted spokes from the central 5-
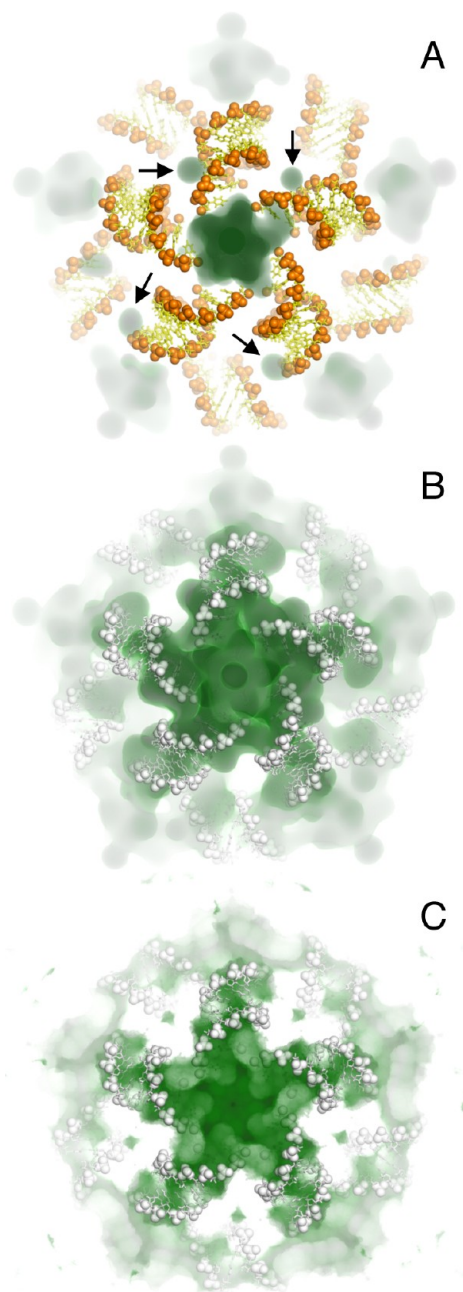
Figure 3. The chloride distribution in STMV relative to the RNA. Regions with high average chloride ion concentration ($>4 \times 10^{-4}$ Å^{-3}) are colored green. Crystallographic RNA is shown for reference in panels B and C. Views along the central icosahedral 5-fold axis seen from the outside of the capsid. (A) MD simulation with RNA. Arrows are explained in the text. (B) MD simulation without RNA. (C). PBE calculation without RNA.

fold axis (Figure 3B). Figure 6A shows the correlation between the chloride ion density in the RNA-free capsid and the crystallographic RNA. Regions with a high presence of RNA in the crystal are densely (zone II) rather than sparsely (zone III) populated by chloride in the empty capsid. Zone I in Figure 6A represents regions with high chloride density in the simulations of the empty capsid that are not populated by RNA in the crystal. These volumes could therefore harbor the unresolved RNA.

The strongest chloride density feature in the RNA-containing simulation was found below the capsid surface on the central 5-fold icosahedral axis as well as at the five equivalent sites closer to the rim (circle a in Figure 1C and Figure 3A). In the electron density maps of the original structure, no significant electron densities above the level of noise could be seen in this region. This region is therefore probably occupied by RNA that does not have icosahedral symmetry in the native virion, e.g., single stranded loops connecting the helices or hairpin loops that fold back onto the same helix. Calculation of the radial distribution of chloride ions around the residues of the protein integrated over the first solvation layer indicate that residues Ser27, Lys30, and Ser121 could potentially be binding RNA (Figure 7). These residues are positioned close to the visible RNA in the direction toward the 5-fold axis and could coordinate the extension of the RNA helix.

Remarkably, the position of the missing nucleotide fragment that was not included in the simulations could be recovered in the chloride ion distributions (marked with arrows in Figure 3A), although not all equivalent sites showed strong chloride ion densities. The Lys11 side chain played a key role coordinating the chloride at this site, and the occupancy depended on the orientation of the flexible N-termini (Figure 7).

The sodium ions in the empty capsid showed a weak tendency of associating with the exterior surface of the capsid, but without any tendency of specific binding. In the simulation with the RNA fragments, one or a few sodium ions were constantly located in the minor groove (i.e., the larger groove in the A type-like double helix) at any given time (arrow b in Figure 1C) much like what has previously been reported for simulations of duplex DNA.[29]

The crystal structure of STMV contains a bound anion on the icosahedral 5-fold axis, possibly a sulfate. The unclear identity of this ion made us decide not to include any anion at these positions. All six equivalent binding sites were constantly occupied by a chloride ion coordinated by the Asn115 and Asn117 side chains (arrow c in Figure 1F, Figure 7A).

Satellite Tobacco Necrosis Virus. In the crystal structure of STNV, a lower amount of RNA was observed than in STMV and the densities are significantly less well-defined. The original crystal structure of STNV did not reveal any densities that could be attributed to packaged RNA.[30−32] Nucleotides could, however, be resolved by using very low resolution data (6−140 Å) in measurements on virus-like-particles (VLP) of the STNV coat proteins containing unidentified, nongenomic RNA. These measurements revealed RNA fragments close to the interior wall of the capsid, in a space constituted by the N-terminal arms of the proteins, which form bundles of three α-helices that extend inward. The RNA fragments could be modeled as short A-form double-stranded helices of three nucleotide pairs with an additional nucleotide stacked onto one of the strands (as can be seen in Figure 4).[14] Low-resolution neutron scattering using heavy water for enhanced contrast had previously revealed cylindrical RNA densities in the space between the N-terminal helix bundles, as well as spherical densities on the 5-fold axes close to the capsid surface.[28]

The N-terminal arm of the STNV CP has been proposed to function as an RNA binding domain,[14,28,30−32] and this is supported by the chloride ion distributions of our simulations. High densities were found around the N-terminal domains (Figure 2), mainly in the clefts between N-terminal helix bundles at the icosahedral 2-fold axes (Figure 4). We also found
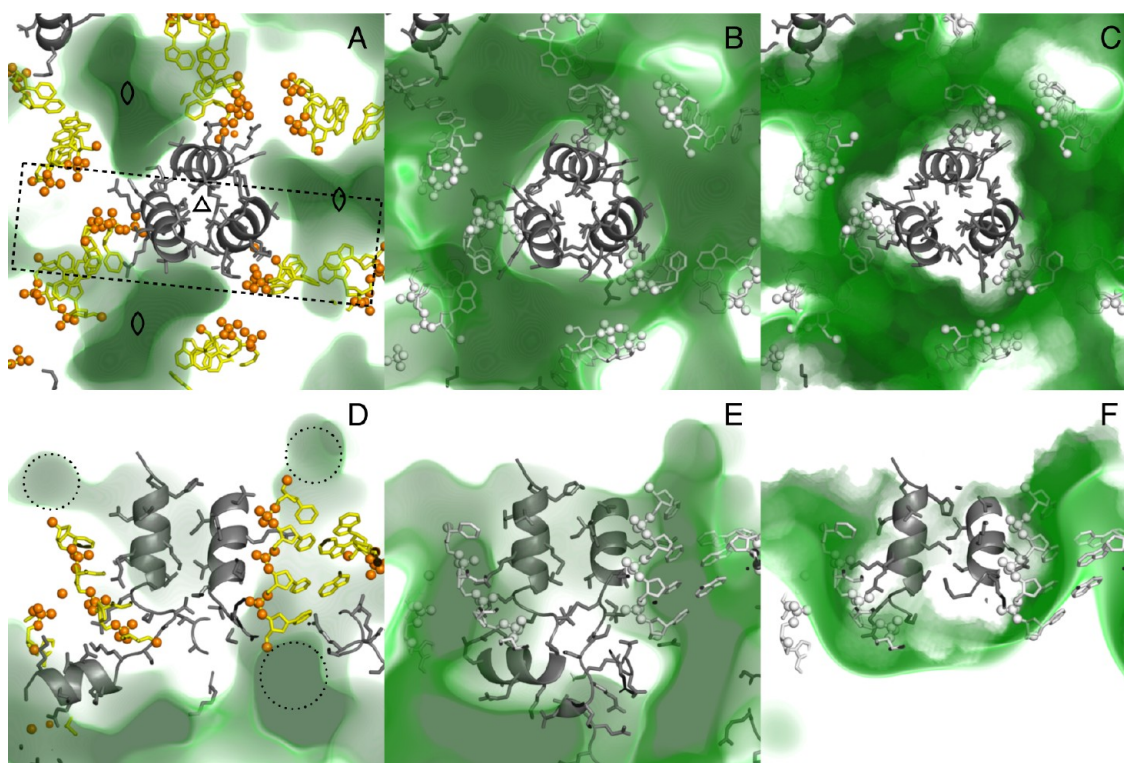
Figure 4. The chloride distribution in STNV around the three α-helix bundle on the 3-fold axis. Regions with high average chloride ion concentration ($>4 \times 10^{-4}$ Å^{-3}) are colored green. Crystallographic RNA are shown for reference in pamels B, C, E, and F. (A−C) Views along an icosahedral 3-fold axis seen from the outside of the capsid. (D−F) Views perpendicular to the views in panels A−C (section marked with a rectangle in A). Up is outside and down is inside the capsid. (A, D) MD simulation with RNA. (B, E). MD simulation without RNA. (C, F) PBE calculation without RNA. (A) The icosahedral 3-fold and 2-fold axes are marked with a triangle and with ovals, respectively. (D) Dashed circles indicate possible additional stacked nucleotides.
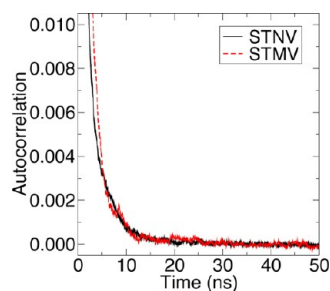

Figure 5. Autocorrelation for chloride ions bound to positively charged protein side chains in the MD simulations of the capsids with no RNA.

a density layer interior of the helices at about 30−35 Å distance from the center of the native virion (Figure 2C). This finding is similar to results from coarse-grained simulations of the RNA distribution in MS2, where the RNA particles display a bimodal radial distribution in the particle attributed to the self-avoidance of the charged RNA.[27]

The crystal structure contains two double helices in each cleft between three-helix bundles of N-terminal domains of the coat protein, in total 60 double helices accounting for at most 420 nucleotides: the height of the RNA electron densities relative to the protein electron densities suggests that not all sites are occupied.[14] The four-nucleotide strand runs parallel and very close to one of the α-helices of the three-helix bundle, while the three-nucleotide strand faces the bundle across the cleft (see Figure 4 for the position of the helices). The chloride ion distribution in the simulation of the RNA-depleted half-capsid

encompasses the entire four-nucleotide strand and most of the three-nucleotide strand (Figure 4B,E). The chloride densities extend further beyond the four-nucleotide strand, suggesting that the protein structure can support an even longer strand of six or seven nucleotides (circled regions in Figure 4D). The simulation with RNA shows chloride ion densities in the region between the two double helices at the 2-fold axes, suggesting that this space can harbor more RNA than what is visible in the crystal structure (Figure 4A).

The side chains of the positively charged residues in the α-helix bundle form salt bridges with the phosphate groups of the RNA fragments. Arg18 binds to the top of the four-nucleotide strand while Arg14 binds to the three-nucleotide strand on the neighboring double helix (Figure 7). The 11 endmost residues were unstructured and did not show any tendency to form a distinct fold but interacted with RNA strands from nearby sites.

Like in the simulations of STMV, a significant chloride ion density was found in the volume just beneath the central 5-fold axis (circle c in Figure 2F) and at all five equivalent sites closer to the half-shell rim. No density was observed here in the crystal structure;[14] however, it was detected in the neutron scattering experiments.[28] The density could harbor hairpin loops or loops joining different helices around the 5-fold axis. Noticeably, this density is not connected with the densities that can be found in conjunction with the N-terminal helix bundles (Figure 2C). The density on the 5-fold axis and the densities around the five N-terminal bundles around the 5-fold axis formed a closed-off region with very low chloride density.

Potential Calcium Binding Site on the Outside of STNV. The capsid of STNV contains 92 calcium binding sites:
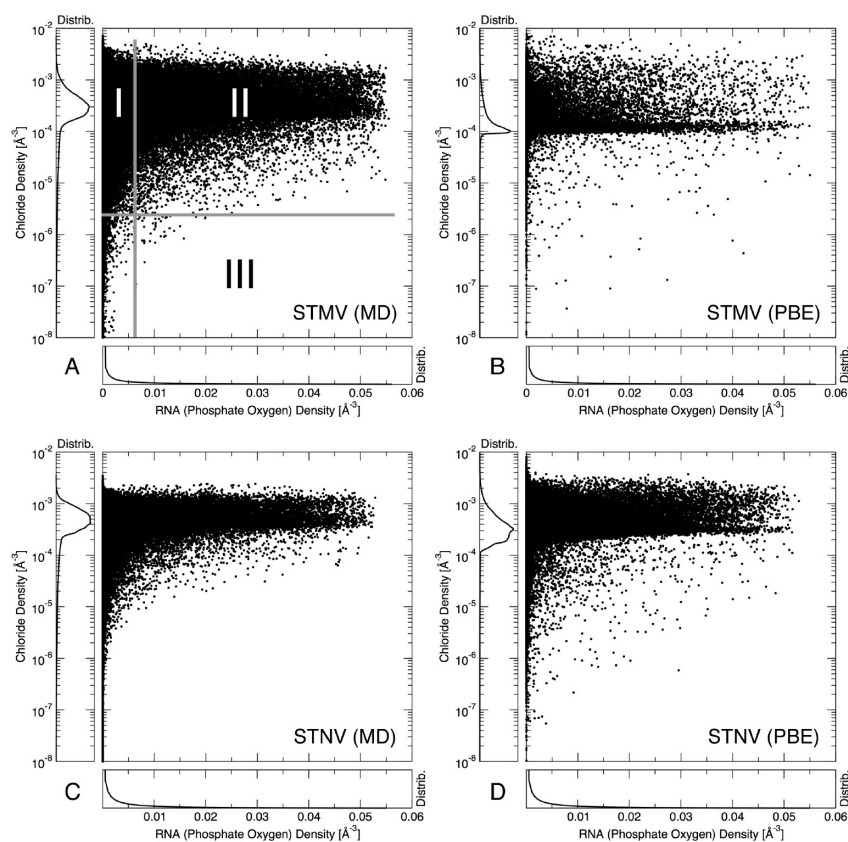
Figure 6. Correlation between chloride ion densities in the empty capsids versus the location of the ordered RNA. The RNA is represented by the distribution of oxygen atoms of RNA phosphate groups in the simulations of the capsids with RNA. (A) Chloride density in STMV from MD. Roman numbers are discussed in the text. (B) Chloride density in STMV from PBE. (C) Chloride density in STNV from MD. (D) Chloride density in STNV from PBE.
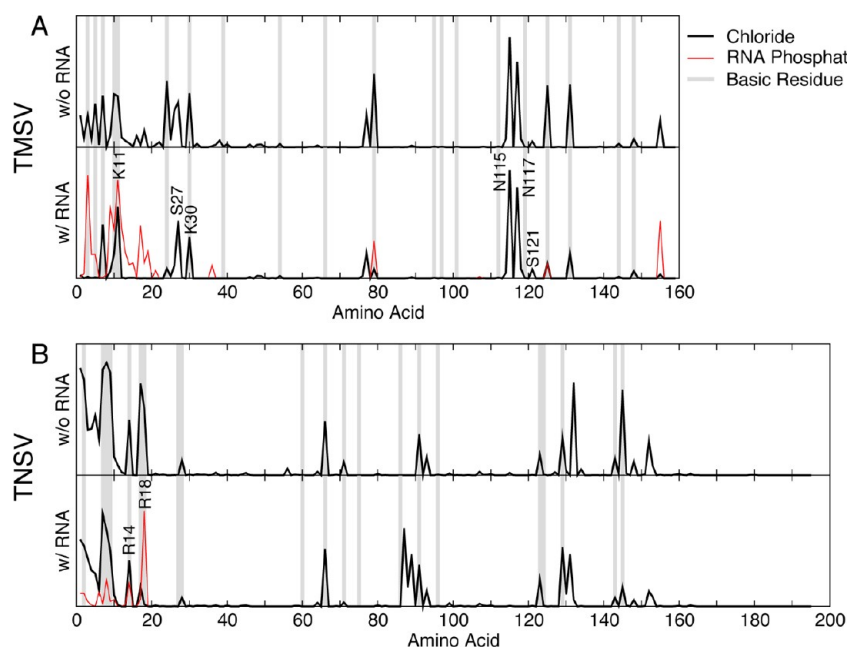

Figure 7. Contacts between protein residues and chloride ions and oxygen atoms of RNA phosphate groups: (A) STMV and (B) STNV.

12 on the 5-fold axis, 20 on the 3-fold axis, and 60 on general positions close to the 3-fold axis. The simulations of STNV suggested that an additional cation binding site is present on the exterior surface of the capsid close to the 3-fold axis consisting of the side chains of Asp56 and Asp63. There are 20

such sites in the half capsid, and they all strongly bound sodium (arrows a in Figure 2D,F) or calcium ions (arrows b in Figure 2A,D) in our simulations. We have previously shown that the four canonical calcium binding sites at the 3-fold axis are the most important ones in controlling the swelling behavior of the

virus and potentially initiate the disassembly of the virus capsid upon entering the host cell.[33] The close proximity to these sites suggests that these binding sites might function as auxiliary calcium binding sites that "attract" and temporarily bind calcium ions to effectively increase the local concentration of calcium in environments with a low ambient concentration of free calcium.

Ion Density from Poisson–Boltzmann Calculations. The electrostatic potential in the volume around the crystal structures of the two empty virus half-capsids was calculated numerically by solving the nonlinear Poisson–Boltzmann (PBE) equation. The calculations included mobile Na^+ and Cl^- ions at a concentration of 150 mM to mimic the MD simulation. The distribution of chloride ions shared many features with the MD simulations close to the protein surface (Figures 3C and 4C,F). However, the PBE calculations could not capture the depth of the chloride ion distributions into the capsid, which can be seen in the molecular dynamics simulations, and it predicted much higher densities close to the protein surface (Figures 8 and 9).
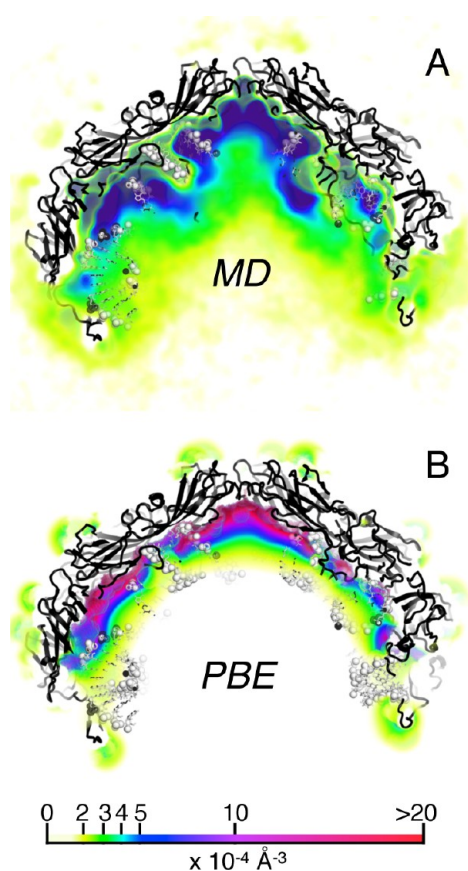

Figure 8. Comparison of chloride distribution of MD and PBE calculations in empty STMV half-capsids. Same view as in Figure 1F. Crystallographic RNA is shown for reference in white. Proteins are rendered in black. (A) MD simulation. (B) PBE calculation.

The correlation plot of the chloride ion density versus the RNA distribution for the PBE calculations (Figure 6B,D) resembles the MD calculations (Figure 6B,C), but shows more noise in the lower right zone of the graph representing false negatives (i.e., low chloride ion density in regions where the crystallographic RNA binds).
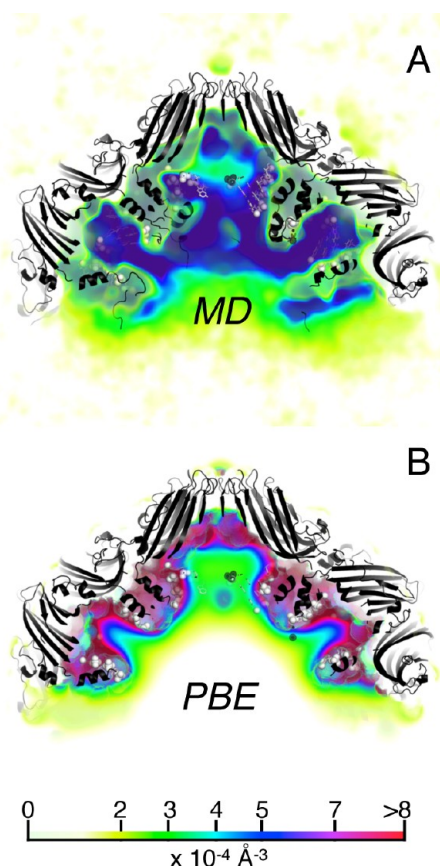

Figure 9. Comparison of chloride distribution of MD and PBE calculations in empty STNV half-capsids. Same view as in Figure 2F. Crystallographic RNA is shown for reference in white. Proteins are rendered in black. (A) MD simulation. (B) PBE calculation.

The STMV map of the PBE calculations contained low densities in a small region on the 3-fold axes. The crystal structure models Asn16 side chains at this location, and due to the symmetry, three of them coincide in a small volume. The equilibrated MD simulation did not display 3-fold symmetry between these three groups, indicating that it could be an artifact of the forced icosahedral symmetry of the crystal structure.

■ DISCUSSION

The genome packaging paradigms are known to vary significantly between different classes of viruses. There is a large discrepancy between helical and icosahedral viruses and also between the double-stranded and the single-stranded viruses.[34] The stiffness of the double-helix makes it difficult to package in a small volume. The capsids of double-stranded DNA viruses assemble first and then molecular motors pump the genome into the capsid container. The structure of the DNA in these virions have been studied using cryo-EM, AFM, and computations. They do not seem to have a unique fold but rather form a coiled structure.[34,35]

The nucleic acids of single-stranded viruses do, on the other hand, in many cases play an integral part in both the assembly of the virion and in maintaining the structural integrity of the virus particle. The question is whether the RNA or DNA molecule or molecules in these viruses have unique folds or not. The averaging required by available imaging techniques makes it impossible to differentiate between these two cases. The

known virus genomes do not have symmetry in the primary sequence, making it impossible to fold them with true icosahedral symmetry. But symmetry on the secondary structure level allows for symmetry at low resolution, for instance at the level of the phosphate backbone. This is indeed the case for a significant fraction of the PaV genome.[11] In addition, secondary structure predictions of the STNV genome do suggest there may be a significant amount of regular RNA structure,[36] in line with the very symmetric distributions of chloride ions found in this work (Figures 2 and 4). On the other hand, many viruses show a high degree of promiscuity toward including foreign RNA or DNA, which would be less likely if the nucleic acid must form a specific secondary structure to fit inside of the capsid.

The RNA and DNA fragments visible in the available virus structures (Table 1) many times consist of secondary structure elements binding to the interior structures of the capsid. These regions of local order might play different roles, such as a packaging signal for the capsid to recognize "its own" nucleic acid. This is most likely the case with MS2, and it has been suggested that a three base-pair helix with a four to six nucleotide long loop with a degenerate AXXA-motif might be a packaging signal in STNV.[36] In large capsids ($T > 1$) biochemically identical CPs must assume different conformations depending on their position in the quasiequivalent symmetry.[37] In the case of FHV it has been proposed that the RNA functions as a molecular switch to induce the structural polymorphism.[10]

Coarse-grained simulations of RNA molecules inside of static protein capsids have hinted that the electrostatics together with the indentations of the protein layer contains a lot of information about the RNA distribution.[26,27] These studies show that virus capsids encode for an environment that can harbor the RNA molecule. The electrostatic properties of the proteins and their relative orientations provide a perfect mold for the RNA molecule to dock into, as is clearly visualized in the chloride distributions inside the STMV (Figures 1 and 3) and STNV (Figures 2 and 4) capsids.

The localization in both capsids of distinct features with high chloride density at the icosahedral 5-fold axis in the region directly interior to the protein capsid suggests that there must be RNA bound here in the native capsids. The high multiplicity at the 5-fold symmetry axis would impose strong constraints on the RNA structure if it were to follow the symmetry. The difficulty in fulfilling these constraints is likely the reason why the viruses do not show signs of ordered RNA there. Instead, these regions probably contain stretches of RNA that link the helical structures together. Whether these links form well-defined structures or flexible loops cannot be assessed with our chloride screening method, however.

The height of the densities at the 5-fold axis are comparable to the densities in the regions where the ordered RNA can be found in the crystal. Therefore, the chloride screening method is not very sensitive to the amount of order in the RNA. Since the method predicts binding locations induced by the protein, the predictive power may be restricted in large cavities of a capsid.

The importance of electrostatic interactions to guide the binding of RNA makes PBE a potentially attractive alternative method to calculate the chloride ion distribution due to the lower computational cost of such calculations (although large-scale PBE calculations are very memory-intensive). Our results show that the PBE maps share features with the MD maps, but

they are not able to capture the depth of the ion distribution inside the capsid. Part of the reason is that the crystal structure used for the PBE calculations did not contain the N-terminal peptides and did not capture the dynamic relaxation of the protein side chains, both of which result in smearing the ion density. Furthermore, these electrostatics calculations treat the solvent as a continuum with a fixed dielectric constant, thereby ignoring the discrete structure of water molecules. The PBE model of the mobile ions used here has the limitation that it does not consider the finite size of the ions, which induces very high ion densities at sites where the potential has high magnitude, e.g., close to the protein surface. A solution to this problem that has been suggested is to limit the maximum local concentration of each ion species.[38]

We have recently reported microsecond simulations of the STNV capsid without RNA.[33] On the basis of those simulations, we were able to pinpoint the locations in the capsid where the virus opens up when the calcium ions are removed. An overall swelling of the capsid was observed, even when the calcium ions were present, which was attributed to the lack of the RNA molecule. Here we presented an elegant method to localize the positions for the RNA inside the entire capsid volume by screening the volume of the capsid interior for RNA binding sites using the distribution of chloride ions. The chloride screening simulations of half a capsid for both STNV and STMV accurately recapitulate the position of the RNA fragments visible in the crystal structures, corroborating the accuracy of the method. The simulations indicate that additional RNA is located close to the capsid surface on the 5-fold axis for both viruses and for STNV also in a layer at about 33 Å from the particle center. These results can aid the construction of atomistic models of entire virus genomes.

■ METHODS

All simulations were performed using the Gromacs 4.5.3 simulation package[39] on a Cray XE6 computer. The simulations were run in parallel using 2016 CPU cores.

System Preparations. Half-capsids of satellite tobacco mosaic virus (PDB structure 1A34[13]) and satellite tobacco necrosis virus (PDB structure 3S4G[14]) were generated from the symmetry matrices within the PDB files. The N-terminal residues for STMV that were not included in the crystal structures (sequence MGRGKVKPNRKS) were modeled as generic loops in clash-free orientation. For STNV (sequence AKQQNNRRKSA) they were modeled as an α-helix pointing away from the three-helix bundle. The helical fold of the N-terminus was chosen because of its compact structure, which facilitated the construction of the peptide in the symmetry-constrained center volume.

Each system contained 30 coat proteins around a central icosahedral 5-fold symmetry axis. Separate systems were generated where the RNA fragments in contact with these proteins were included. The 5′ phosphate groups of the RNA fragments were not included and neither was the nucleotide fragment found on a general location in the STMV structure. Each of the calcium binding sites in STNV structure was occupied by a calcium ion (from the PDB structure 2BUK[32]). The systems were solvated in water in rhombic dodecahedron boxes with sides of about 23 nm length (the outer diameters of the two capsids are 17.6 and 19.8 nm, respectively) and with periodic boundary conditions. Random water molecules were replaced with sodium and chloride ions until neutral systems were obtained at an ionic strength of about 150 mM (Table 2).

Simulation Scheme and Parameters. Each system was energy minimized using the steepest gradient method for 1000 steps, sufficient for convergence of the maximum force on any atom. The systems were then simulated with position restraints applied on all non-hydrogen protein and RNA atoms for 1 ns to relax the solvent and then for another 20 ns without the position restraints on the modeled N-termini to let them relax and to equilibrate the pressure of the simulation box. Finally, the systems were subject to simulations for 200 ns. During the latter half of these 200 ns, the position restraints were only applied to the C_α-atoms of the proteins (excluding the modeled N-termini) and the C1′-atoms of the ribose ring of the RNA.

The Amber99sb-ILDN force field[40,41] and the TIP3P water model[42] were used with an integration time-step of 2 fs. The solvent molecules were constrained with the SETTLE algorithm[43] and the bond lengths of the macromolecules were constrained with the parallel LINCS algorithm.[44] Short range Coulomb and van der Waals (Lennard-Jones) interactions were calculated within a cutoff of 1.15 nm. Neighbor lists were used and these were updated every 10th step. Long range electrostatic interactions were calculated with the particle-mesh Ewald method[45] with a Fourier grid spacing of 0.130 nm and PME-order (Figure 4). Long-range dispersion corrections were applied in the calculation of the energy and the pressure. The temperatures of the macromolecules and the solvent were coupled separately to an external heat-bath at 300 K using the V-rescale algorithm[46] with a coupling time constant of 0.5 ps. The average pressure of the simulation box was kept constant at 1 bar, initially using the Berendsen algorithm[47] with a coupling time constant of 25 ps and compressibility coefficient of 4.5×10^{-5} bar^{-1} and later, during the final 200 ns of simulation, using the Parrinello–Rahman algorithm[48,49] with a time constant of 5 ps.

Poisson–Boltzmann Calculations. The nonlinear Poisson–Boltzmann equation (PBE) was solved numerically for the crystal structures of the two virus half-capsids using the Adaptive Poisson–Boltzmann Solver (APBS) software.[50] The experimentally unresolved N-terminal peptides were not included in the PBE model. The calculations were performed using a multigrid approach[51,52] with a fine grid spacing of 0.5 Å. A molecule surface was used for mapping the dielectric constant and ion accessibility. It was constructed by using a spherical solvent-molecule probe with a radius of 1.4 Å. The point charges were spread on the grid using trilinear interpolation. The potential at the boundary was estimated using the single Debye–Hückel method. Dielectric constants of 2 (inside the proteins) and 78.54 (elsewhere) were used. Mobile monovalent cations and anions at a 150 mM concentration were used. The temperature was 298.15 K. All other parameters were set to the default value or the value recommended in the APBS user guide. The calculations were done in sequence, a single calculation requiring about 10 GB of memory.

Analysis. The time-averaged distributions of sodium and chloride ions as well as RNA phosphate oxygens of the MD simulations were converted into density maps using the VolMap tool of the VMD molecular visualization program (version 1.9).[53] The maps were constructed from three-dimensional Gaussian functions placed at the positions of the ion/atom from each frame in the final 100 ns (in total 10 001 frames per map). The width (standard deviation) of the Gaussian was taken as twice the van der Waals radius of the ion/atom in question as defined by VMD. Difference maps

were calculated as the sodium map minus the chloride map for both viruses. The difference maps were used for the rendition of Figures 1 and 2. The overlaps between the two ion species were small.

The chloride ion concentrations in the PBE calculations were calculated from the electrostatic potential $\Phi(x,y,z)$ on the grid as

$$C(x, y, z) = C_0 \exp \frac{\Phi(x, y, z)}{k_B T}$$

where C_0 is the background ion concentration (150 mM).

Chloride ions were considered to be bound to the protein if they were located within a cutoff of 0.4 nm around the side chain nitrogen atoms in lysine or arginine. This definition yields a binary function of time for each pair of protein side chains and ions. The autocorrelation of this function was averaged over all pairs (Figure 5) and the last 100 ns of the simulations. The integral of this function yields a rough lifetime for ion binding; although this definition may not in all cases yield the most realistic lifetime,[54] it is sufficient for the purpose of demonstrating that the ions are in equilibrium with the protein structure.

The correlation plots of the chloride versus RNA density were constructed by plotting the chloride ion concentration versus the RNA phosphate oxygen concentration where the oxygen density was nonzero. In Figure 5 every 10th point is shown for increased clarity. In addition Figure 5 also shows the distribution of concentrations of the gridpoints in the maps along both axes.

Protein–ion and protein–RNA contacts were evaluated by calculating the radial distribution function (RDF) of solvent chloride ions and RNA phosphate group oxygens around the oxygen and nitrogen groups of each residue in the five protein chains around the central 5-fold icosahedral symmetry axis. The first peak of the RDF averaged over the 5-fold symmetry was integrated for each residue to obtain the number of contacts. The normalized results are plotted versus the protein sequence in Figure 7.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: spoel@xray.bmc.uu.se.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Neutze, R.; Wouts, R.; van der Spoel, D.; Weckert, E.; Hajdu, J. *Nature* **2000**, *406*, 752−757.

(2) Maia, F. R. N. C.; Ekeberg, T.; Timneanu, N.; van der Spoel, D.; Hajdu, J. *Phys. Rev. E* **2009**, *80*, 031905.

(3) Wang, F.; Weckert, E.; Ziaja, B.; Larsson, D. S. D.; van der Spoel, D. *Phys. Rev. E* **2011**, *83*, 031907.

(4) Chen, Z.; Stauffacher, C.; Li, Y.; Schmidt, T.; Bomu, W.; Kamer, G.; Shanks, M.; Lomonossoff, G.; Johnson, J. *Science* **1989**, *245*, 154−159.

(5) Lin, T.; Cavarelli, J.; Johnson, J. E. *Virology* **2003**, *314*, 26−33.

(6) Wery, J.-P.; Reddy, V. S.; Hosur, M.; Johnson, J. E. *J. Mol. Biol.* **1994**, *235*, 565−586.

(7) Speir, J. A.; Munshi, S.; Wang, G.; Baker, T. S.; Johnson, J. E. *Structure* **1995**, *3*, 63−78.

(8) Larson, S. B.; Day, J.; Canady, M. A.; Greenwood, A.; McPherson, A. *J. Mol. Biol.* **2000**, *301*, 625−642.

(9) Valegård, K.; Murray, J. B.; Stonehouse, N. J.; van den Worm, S.; Stockley, P. G.; Liljas, L. *J. Mol. Biol.* **1997**, *270*, 724−738.

(10) Fisher, A. J.; Johnson, J. E. *Nature* **1993**, *361*, 176−179.

(11) Tang, L.; Johnson, K. N.; Ball, L. A.; Lin, T.; Yeager, M.; Johnson, J. E. *Nat. Struct. Biol.* **2001**, *8*, 77−83.

(12) Speir, J. A.; Taylor, D. J.; Natarajan, P.; Pringle, F. M.; Ball, L. A.; Johnson, J. E. *Structure* **2010**, *18*, 700−709.

(13) Larson, S. B.; Day, J.; Greenwood, A.; McPherson, A. *J. Mol. Biol.* **1998**, *277*, 37−59.

(14) Lane, S. W.; Dennis, C. A.; Lane, C. L.; Trinh, C. H.; Rizkallah, P. J.; Stockley, P. G.; Phillips, S. E. V. *J. Mol. Biol.* **2011**, *413*, 41−50.

(15) Lucas, R. W.; Larson, S. B.; Canady, M. A.; McPherson, A. *J. Struct. Biol.* **2002**, *139*, 90−102.

(16) Larson, S. B.; Lucas, R. W.; Greenwood, A.; McPherson, A. *Virology* **2005**, *334*, 245−254.

(17) Tsao, J.; Chapman, M.; Agbandje, M.; Keller, W.; Smith, K.; Wu, H.; Luo, M.; Smith, T.; Rossmann, M.; Compans, R.; Parrish, C. *Science* **1991**, *251*, 1456−1464.

(18) Llamas-Saiz, A. L.; Agbandje-Mckenna, M.; Parker, J. S.; Wahid, A.; Parrish, C. R.; Rossmann, M. G. *Virology* **1996**, *225*, 65−71.

(19) Bernal, R. A.; Hafenstein, S.; Olson, N. H.; Bowman, V. D.; Chipman, P. R.; Baker, T. S.; Fane, B. A.; Rossmann, M. G. *J. Mol. Biol.* **2003**, *325*, 11−24.

(20) McKenna, R.; Xia, D.; Willingmann, P.; IIag, L. L.; Krishnaswamy, S.; Rossmann, M. G.; Olson, N. H.; Baker, T. S.; Incardona, N. L. *Nature* **1992**, *355*, 137−143.

(21) Bernal, R. A.; Hafenstein, S.; Esmeralda, R.; Fane, B. A.; Rossmann, M. G. *J. Mol. Biol.* **2004**, *337*, 1109−1122.

(22) LlamasSaiz, A.; AgbandjeMcKenna, M.; Wikoff, W.; Bratton, J.; Tattersall, P.; Rossmann, M. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1997**, *53*, 93−102.

(23) Kontou, M.; Govindasamy, L.; Nam, H.-J.; Bryant, N.; Llamas-Saiz, A. L.; Foces-Foces, C.; Hernando, E.; Rubio, M.-P.; McKenna, R.; Almendral, J. M.; Agbandje-McKenna, M. *J. Virol.* **2005**, *79*, 10931−10943.

(24) van der Graaf, M.; Scheek, R. M.; van der Linden, C. C.; Hemminga, M. A. *Biochemistry* **1992**, *31*, 9177−9182.

(25) van der Spoel, D.; Feenstra, K. A.; Hemminga, M. A.; Berendsen, H. J. C. *Biophys. J.* **1996**, *71*, 2920−2932.

(26) Zhang, D.; Konecny, R.; Baker, N. A.; McCammon, J. A. *Biopolymers* **2004**, *75*, 325−337.

(27) ElSawy, K. M.; Caves, L. S. D.; Twarock, R. *Biophys. J.* **2011**, *101*, 774−780.

(28) Bentley, G.; Lewit-Bentley, A.; Liljas, L.; Skoglund, U.; Roth, M.; Unge, T. *J. Mol. Biol.* **1987**, *194*, 129−141.

(29) Rueda, M.; Cubero, E.; Laughton, C.; Orozco, M. *Biophys. J.* **2004**, *87*, 800−11.

(30) Unge, T.; Liljas, L.; Strandberg, B.; Vaara, I.; Kannan, K. K.; Fridborg, K.; Nordman, C. E.; Lentz, P. J. *Nature* **1980**, *285*, 373−377.

(31) Liljas, L.; Unge, T.; Jones, T.; Fridborg, K.; Lövgren, S.; Skoglund, U.; Strandberg, B. *J. Mol. Biol.* **1982**, *159*, 93−108.

(32) Jones, T.; Liljas, L. *J. Mol. Biol.* **1984**, *177*, 735−767.

(33) Larsson, D. S. D.; Liljas, L.; van der Spoel, D. *PLoS Comput. Biol.* **2012**, *8*, e1002502.

(34) Harvey, S. C.; Petrov, A. S.; Devkota, B.; Boz, M. B. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10553−10564.

(35) Purohit, P.; Kondev, J.; Phillips, R. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3173−3178.

(36) Bunka, D. H. J.; Lane, S. W.; Lane, C. L.; Dykeman, E. C.; Ford, R. J.; Barker, A. M.; Twarock, R.; Phillips, S. E. V.; Stockley, P. G. *J. Mol. Biol.* **2011**, *413*, 51−65.

(37) Caspar, D. L. D.; Klug, A. *Cold Spring Harb. Symp. Quant. Biol.* **1962**, *27*, 1−24.

(38) Gruziel, M.; Grochowski, P.; Trylska, J. *J. Comput. Chem.* **2008**, *29*, 1970−1981.

(39) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.

(40) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712−725.

(41) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950−1958.

(42) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(43) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952−962.

(44) Hess, B. *J. Chem. Theory Comput.* **2008**, *4*, 116−122.

(45) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.

(46) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

(47) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(48) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182−7190.

(49) Nosé, S.; Klein, M. L. *Mol. Phys.* **1983**, *50*, 1055−1076.

(50) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037−10041.

(51) Holst, M.; Saied, F. *J. Comput. Chem.* **1993**, *14*, 105−113.

(52) Holst, M. J.; Saied, F. *J. Comput. Chem.* **1995**, *16*, 337−364.

(53) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.

(54) van der Spoel, D.; van Maaren, P. J.; Larsson, P.; Tîmneanu, N. *J. Phys. Chem. B.* **2006**, *110*, 4393−4398.