A Complete Virion Simulated: All-Atom Model of MS2 Bacteriophage with Native Genome

Vladimir S. Farafonov,[a,b] Michael Stich[c,b] and Dmitry Nerukh*[b]*

[a] Department of Physical Chemistry, V. N. Karazin Kharkiv National University, Kharkiv, 61022, Ukraine
[b] Department of Mathematics, Aston University, Birmingham, B4 7ET, UK
[c] Área de Matemática Aplicada, Departamento de Matemática Aplicada,
Ciencia e Ingeniería de Materiales y Tecnología Electrónica,
Universidad Rey Juan Carlos, C. Tulipán s/n, 28933 Móstoles (Madrid), Spain

Abstract For the first time, a complete all-atom Molecular Dynamics model of a virus, bacteriophage MS2, in its entirety including protein outer shell, native genomic RNA with necessary divalent ions, and surrounding explicit aqueous solution with ions at physiological concentration was built. The model is based on an experimentally measured cryo-EM structure, which was substantially augmented by reconstructing missing or low-resolution parts of the measured density (where the atomistic structure cannot be fit unambiguously). The model was tested by a quarter of a microsecond MD run and various biophysical characteristics are obtained and analysed. The developed methodology of building the model can be used for reconstructing other large biomolecular structures when experimental data is fragmented and/or of varying resolution, while the model itself can be used for studying the biology of MS2, including the dynamics of its interaction with the host bacteria.

1. Introduction

All-atom dynamic models of viruses, even though started to appear nearly two decades ago, are still rare nowadays. In our 2018 perspective [1] we described all publications on the subject at the time and the situation remains largely the same today. While all-atom outer shells, capsids, are modelled for several viruses (see, for example, a recent review [2]), reconstructing the encapsulated genome is still a significant challenge. To the best of our knowledge, there are no reported all-atom structures of a *native* genome in its entirety for any virus, even though a model with artificial RNA of length similar to the native genome for STMV virus [3] as well as an attempt to reconstruct the native genome of MS2 bacteriophage from the experimental structural data [4] were published. We would like to stress that there are many models of genomes and virus particles in general at lower resolution (larger scales), from coarse-grain Molecular Dynamics to representing parts of viruses as elastic bodies, however, *atomistic* reconstructions based on experimental structures are still missing.

Atomistic models of such large biomolecular systems are important, especially in light of the recent developments in cryo-EM measurements that reached unprecedented levels. It is now possible to measure the structure of whole cellular organelles and other biomolecular assemblies of sizes exceeding the size of smallest viruses at the resolution of less than 3Å, which is usually enough for fitting an atomistic model. However, almost always, the resolution of the measured density map varies at different locations of the system, such that even if most of the system is measured at high enough resolution to fit the model, there are (often significant) parts of the system that are measured at lower resolution or not measured at all. In this situation, methods of building *complete* all-atom models and their subsequent validation with dynamics are critically important.

Virus particles are convenient systems for building such models as they are self-contained and stable in solution. They do not require interactions with other parts of a cell, as is often the case with cellular organelles. At the same time, they are complete biological systems that can be studied isolated or in interaction with cellular subsystems. MS2 is a bacteriophage infecting *E. coli* which is used extensively as a model virus due to its small size, simple structure, and relatively straightforward synthesis [5]. This allowed numerous modifications of the outer and inner surfaces of its capsid as well as manipulation of its genome. Structurally, the capsid consists of 89 dimers of one coat protein and one copy of the 'maturation protein', while the genome is represented by a 3569 nucleotides long single-stranded RNA molecule. There are no other molecules in this virus. Dynamically, however, infecting bacteria involves a complicated mechanism that includes the bacterial pilus and a still unknown process of extracting the viral RNA from the capsid and transferring it in the bacteria [6–8]. The capsid itself remains largely intact outside the

bacteria. Therefore, the detailed all-atom dynamic model of MS2 will help to elucidate this mechanism as well as help in various engineering methodologies for this virus.

We have built an almost complete all-atom atomistic model of an empty MS2 capsid [9] that revealed subtle details of its structure and dynamics. We have also reconstructed a complete and operable all-atom model of its genome using mixed-resolution cryo-EM density [10]. As a step in the genome model preparation we assembled a complete MS2 virion in order to fit the RNA model into the capsid and simulated it for 50 ns to check the robustness of the genome model [10–12]. While our previous work was mostly focused on either capsid or genome subsystems separately, we here report the details of the procedure of assembling the whole virion and investigate the biophysical properties of the complete virus particle, computed from an extended (265 ns) MD simulation.

2. Methods

2.1. Genome reconstruction

We used the model from [10] with no additional adjustments. The input data for reconstruction were the experimentally resolved incomplete RNA backbone, the 16 experimentally resolved stem-loops, and the computational prediction of RNA secondary structure (base pairing). Briefly, the process consisted of the following steps:

(i) identifying stem-loops in the secondary structure;
(ii) generating a perfect all-atom structure of each stem-loop;
(iii) manually placing the perfect stem-loops over the resolved backbone;
(iv) ensuring the absence of steric clashes, gaps, overlaps between the arranged stem-loops;
(v) joining all parts together into the single RNA chain;
(vi) energy minimization on the coarse-grained and then all-atom levels of detail;
(vii) fitting the RNA into the capsid to remove steric clashes between the two constituents.

Step (iv) was the central difficulty. The original secondary structure prediction appeared incompatible with the backbone at several regions, so it was necessary to try several variants of the secondary structure and choose the most suitable one.

2.2. Capsid reconstruction

In our capsid model [9] four intervals of the maturation protein (MP) sequence located at its "tip" were omitted because they were not resolved in the experiment. Here, we reused that model and restored the missing intervals (63 residues in total). The amino acid residues of MP before and after the missing parts form β-sheets. We assumed that this motif extends to the missing residues as well. Therefore, four generic β-sheets were prepared from the known amino acid sequence and then manually placed to roughly fill the gaps using Avogadro software [13]. The obtained model of MP was energy minimised in vacuum and in water, and then the new β-sheets were equilibrated in water during 20 ns MD simulation, keeping the experimentally resolved part of MP restrained in the original cryo-EM conformation. The procedures were done in the absence of the capsid. The completed MP was placed to the capsid model instead of the original one, and the position was found using least-squares fitting of their experimentally resolved parts. The complete maturation protein in its place is shown in Fig. 1.

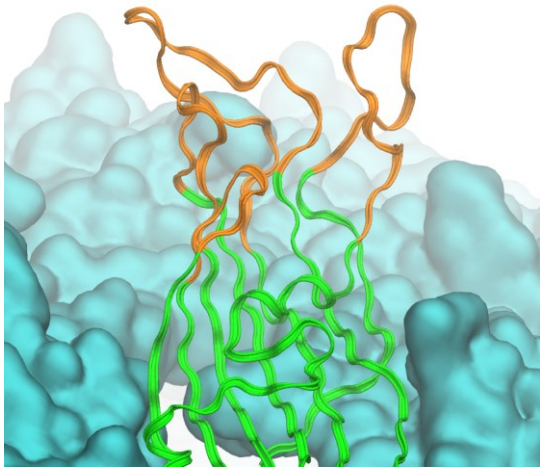


Figure 1. The maturation protein (green) with the reconstructed tip (orange). The rest of the capsid is coloured blue.

2.3. Virion assemblage

The two main constituents of the virion, the genome and its protein envelope, the capsid, were assembled together. The RNA structure presented in [10] was designed to fit the capsid with no steric clashes, therefore it was enough just to superpose the two constituents. The positioning was performed by least-squares fitting of the 16 experimentally resolved stem-loops of the genome to their experimentally determined positions. After this a short MD run was carried out in vacuum to bring these 16 stem-loops closer to their cryo-EM shapes. Pull code with constant pulling force was applied here [14].

2.4. Solvation

The prepared assembly of an approximately spherical shape was solvated in a dodecahedron-shaped box. The edge length was set to 21.4 nm, which allowed for 2×5.5 nm margin around the capsid after the application of three-dimensional periodic boundary conditions.

The assembly is strongly charged and correct neutralisation of this charge with counter-ions is crucial for the stability of the model [15]. The RNA is charged strongly negatively (–3568 *e*) while the capsid wall has a net positive charge (+191 *e*). Importantly, the latter charge is distributed very unevenly between the sides of the capsid, so that the inner wall of the capsid is positively charged (+644 *e*) and the outer wall is negatively charged (–453 *e*) [9]. The virus particles are known to include double charged Mg^{2+} cations which play a role in maintaining the correct genome folding [16]. Therefore, we separately neutralised the capsid interior by randomly placing (3568 – 644)/2 = 1462 Mg^{2+} cations inside it, and the external solution was neutralised by distributing 453 Na^+ cations there. As the final step, Na^+ and Cl^- ions were added at physiological saline concentration (0.15 mol L^{-1}) both to the interior (406 ion pairs) and to the bulk solution (2351 ion pairs). A complete model of an MS2 virion in laboratory conditions was thus obtained, presented in Fig. 2.

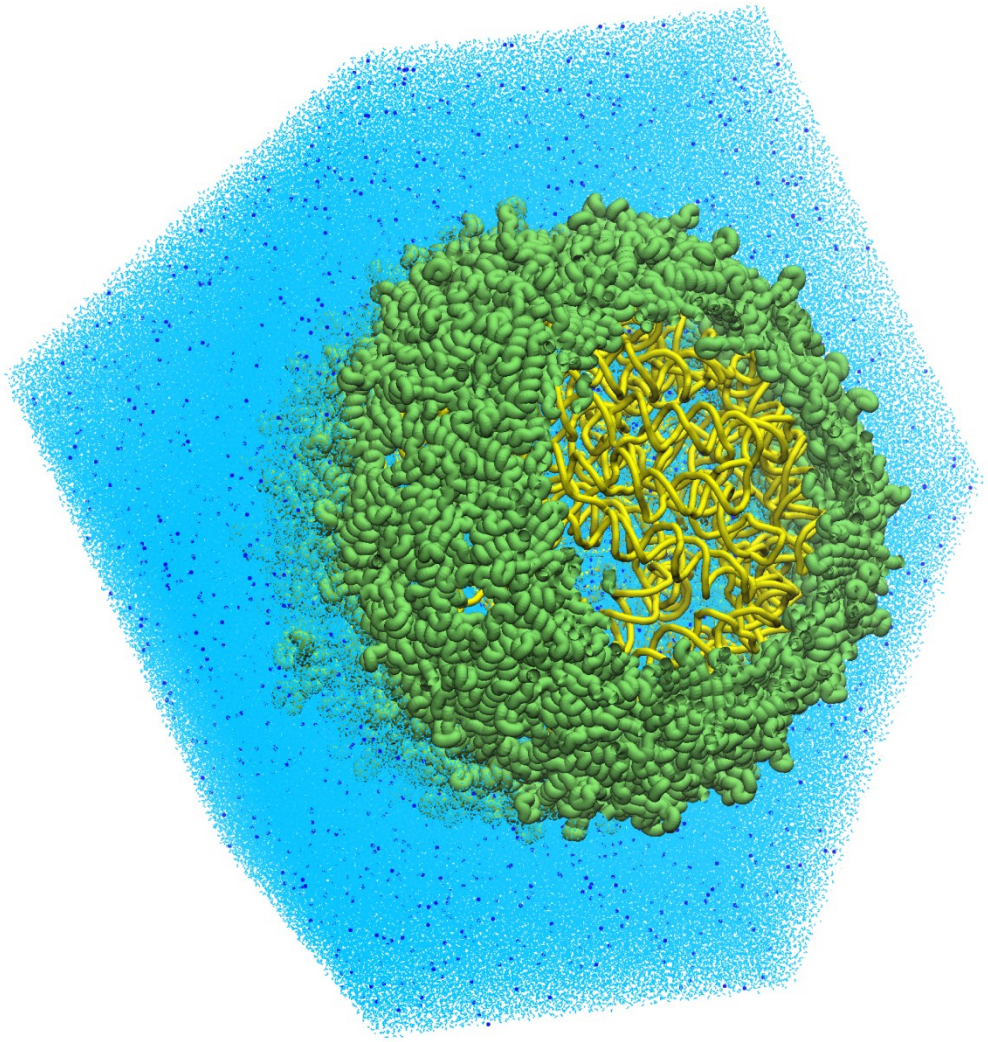


Figure 2. Assembled and solvated model of MS2 virion. Dark blue points represent ions.

2.5. Force field

A proper choice of the force field is vital for producing accurate results in MD simulation. For the virion components we selected a consistent set of force fields, which are recent and at the same time well-tested. The capsid was described with AMBER ff14SB, a recent development in the well-known series of force fields [17]. For the genome we took AMBER ff99bsc0 with χ_{OL3} correction [18] that is recognised to be among the most suitable for RNA simulations [19]. Importantly, these protein and RNA force fields are compatible with each other because both are derived from AMBER ff99. A suitable water model was TIP3P, which motivated us to prefer the ff14SB version of the forcefield to its more recent development ff19SB: the latter is tailored for the OPC water model, which is still uncommon and more computationally expensive.

For Na^+ and Cl^- ions the parameters derived by the Cheatham group were taken [20] because they were derived specifically for simulating solutions of biomolecules with AMBER force field. Representation of the Mg^{2+} cation is difficult due to the very strong electric field around the ion, resulting in strong polarisation ability. We employed the elaborate potential model from Ref. [21]. There, the hydration shell consisting of six water molecules is explicitly attached to the Mg^{2+} cation, and their charge distribution accounts for their polarisation by the ion. Finally, the CUFIX correction [22] that contains several adjusted intra-protein, protein–RNA, ion–ion, ion–protein, and ion–RNA parameters was applied, which was shown to improve the accuracy of biomolecular simulations [23]. A notable feature of this combination is its validation with a system very similar to the one considered here, namely, a dense DNA array dissolved in a concentrated saline [21]. All the above parameters are available as a single package, and we found this option convenient [24].

2.6. Simulation parameters

For simulations, the GROMACS 5 software package was used. The temperature of MD simulations was set to 298 K to conform to laboratory conditions. Three-dimensional periodic boundary conditions were imposed, the time step was equal to 2 fs, and all covalent bonds were constrained by LINCS algorithm. Electrostatic interactions were computed with the PME method, while van der Waals interactions were cut off at 1 nm; dispersion corrections to energy and pressure were evaluated. After energy minimisation with the steepest descent method, the system was prepared in several stages:

(i) 10 ns run in NVT ensemble with both the capsid and genome restrained at their initial positions; the virion and the saline were thermostated separately with the Berendsen thermostat;
(ii) 10 ns run at the same conditions, but only 16 resolved stem-loops were kept restrained, the rest of the virion was unconstrained;
(iii) 60 ns run at the same conditions, but all components were free to move;
(iv) 5 ns run in NPT ensemble; pressure was kept at 1 bar with the Berendsen barostat;
(v) finally, a 200 ns long run was carried out; the velocity-rescale thermostat and Parrinello-Rahman barostat were used instead of the Berendsen couplings to produce a correct NPT ensemble.

The analysis of the structural characteristics allowed to conclude that during the first ca. 100 ns of the run of stage (v) the structure continued to equilibrate. Therefore, as production run, we considered the last 100 ns of the stage (v) simulation. The time averaged quantities presented below are determined using this production run. However, when time evolution of a quantity is presented, the equilibration runs (iii), (iv), and 0–100 ns interval of (v) were also shown.

2.7. Empty capsid simulation

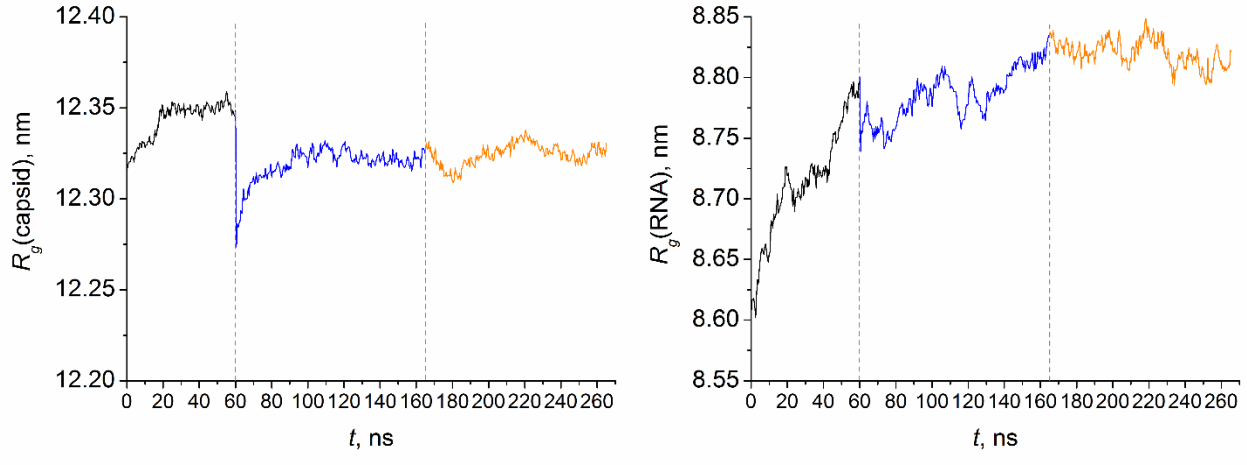
The previously reported simulation of the capsid with incomplete MP was carried out using AMBER03 force field [9]. Therefore, to allow for correct comparison of the empty and genome-loaded capsids we repeated that simulation using AMBER14sb force field and complete maturation protein. The composition of the solution was preserved. The procedure was as follows:

(i) 10 ns run in NVT ensemble with restraints and separate thermostating of protein and saline with the Berendsen thermostat;
(ii) 10 ns run in NPT ensemble without restraints; pressure was kept at 1 bar with the Berendsen barostat;
(iii) 50 ns production run in NPT ensemble with velocity-rescale thermostat and Parrinello-Rahman barostat.

3. Results

3.1. Stability of the model

Simple but informative characteristics of the stability of a modelled biomolecule are the time evolutions of the gyration radius R_g and the root mean square displacement RMSD. For an equilibrated model of a stable assembly, they are expected to fluctuate around some constant average value. The graphs separately for the capsid and the genome are shown in Figs. 3, 4.



a **b**

Figure 3. **a**: Gyration radius of the RNA, **b**: gyration radius of the capsid. The black curve is the NVT equilibration, the blue curve is the NPT equilibration, the orange curve is the production run. Dashed lines indicate the boundaries of the runs.

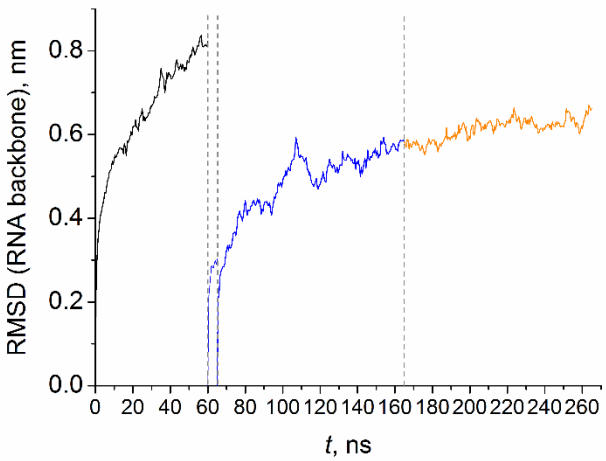


Figure 4. RMSD of the RNA backbone. The black curve is NVT equilibration, the blues curves are two parts of NPT equilibration, the orange curve is the production run. Dashed lines indicate the boundaries of the runs.

We estimated that the gyration radii of the capsid reached their equilibrium value of 12.3 nm at ca. 100 ns, while for RNA the equilibrium value of 8.8 nm settled at ca. 150 ns. RMSD of RNA demonstrated minor increase after 130 ns and reached the constant value at ca. 200 ns.

Fig. 5 shows the time evolution of the number of RNA–capsid and RNA–RNA hydrogen bonds. The former characteristic considerably increased during the equilibration (from 330 to 520 on average) indicating that RNA more tightly fitted the capsid compared to the original reconstructed configuration. The growth of the latter characteristic was more limited (from 4500 to 4610 on average). Both quantities reached the value close to the final value after 140–150 ns, but still were subjected to fluctuations. Based on this information we chose 165 ns as the beginning of the production run and used this interval to calculate the values presented below.

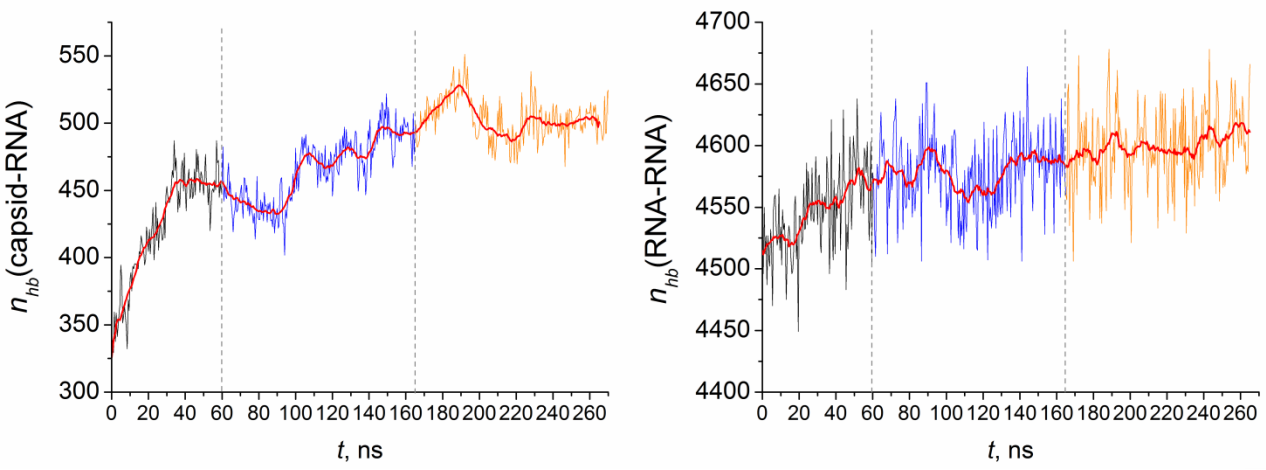


Figure 5. **a**: Number of RNA–capsid hydrogen bonds, **b**: Number of RNA–RNA hydrogen bonds. The black curve is NVT equilibration, the blue curve is NPT equilibration, the orange curve is production run, the red curve is adjacent average over 10 ns. Dashed lines indicate the boundaries of the runs.

3.2. Flexibility of the genome

With the dynamic model built, it is now possible to elucidate the reason for the inhomogeneous character of the cryo-EM density used for reconstructing the genome. The effects that lead to blurring the electron density may be classified into two main categories: conformation disorder, when some portions of the molecule can have different positions among the samples, and thermal motion. The latter also occurs on a wide spatial and temporal scale: from oscillations of individual atoms around their equilibrium positions to collective motions of secondary structure elements and domains.

Here we examined the correspondence between the experimental resolution of the nucleotides and their root mean square fluctuation (RMSF) at a long (100 ns) time scale. The experimental resolution was represented by a number from the set {0, 1, 2}, where 0 (high resolution) was assigned to the 16 atomistically reconstructed stem-loops, 1 (moderate resolution) to the residues placed over the traced backbone, and 2 (low resolution) to the rest of the genome, which was not resolved in the experiment. The RMSF was computed over the RNA backbone atoms using the production run. The values are compared in Figure 6.

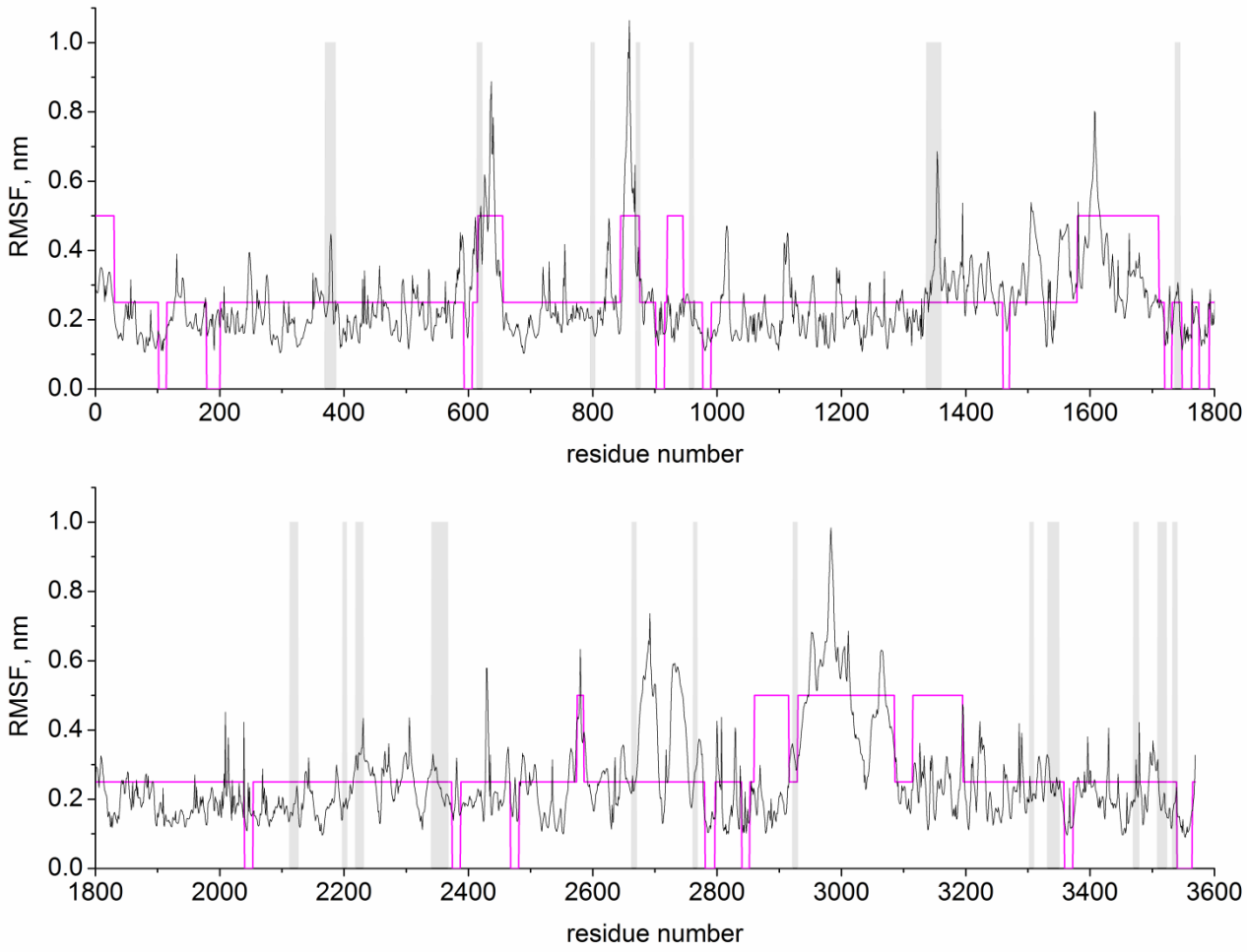


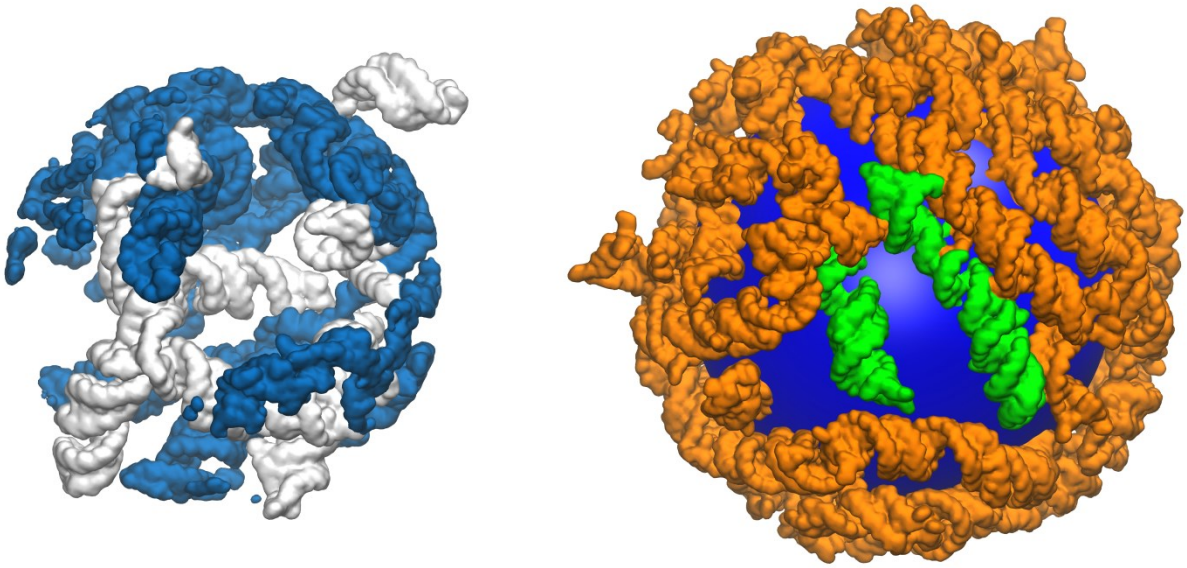
Figure 6. The experimental resolution (violet) and the RMSF from a 100 ns MD simulation (black) of the nucleotides. Greyed are single-stranded intervals longer than 6 residues. The violet curve is scaled down 4 times for visual purpose.

There is a pronounced correspondence between the two metrics. The nucleotides showing RMSF higher than ~0.5 nm are mostly unresolved. Long single stranded chains correspond to local maxima of RMSF in some cases. In contrast, the precisely-resolved nucleotides always correspond to local minima: their RMSF is 0.1–0.15 nm. The cryo-EM density of residues, which express intermediate values of 0.2–0.5 nm, allowed the reconstruction of the backbone only.

The visualisation of the nucleotides without the underlying traced backbone reveals that almost all of them stay deep in the capsid (i.e. contained within 8 nm from its centre), see Fig. 7a. The only exception is the starting stem-loop 1–28 staying beyond 8 nm and the protruding tips of long stem-loops.

Still, there are several intervals which do not fit into this scheme. In the following, we discuss these intervals in more detail. First, the intervals 2860–2915 and 3115–3195 show low RMSF but have no resolved backbone. Visual inspection shows that these are two long stem-loops, which have much free space around them, see Fig. 7b. This is exceptional since most of the other loops are in a crowded environment and are clamped between the surrounding parts of the genome. Hence, the low resolution of the cryo-EM density associated with the two intervals arises not from collective motions (which are weak according to RMSF), but from the conformational heterogeneity: in different viral particles these stem-loops possibly were placed differently within the large available space.

Second, the intervals 1501–1521 & 1547–1569 and 2672–2708 & 2716–2754, as well as several short intervals, have their backbone traced despite considerable RMSF. The four former intervals are parts of very long stem-loops 1496–1573 and 2671–2752, whose ends (1522–1546 and 2709–2715) contact with the capsid and have a low RMSF. Examination shows that these pieces fluctuate like strings stretched between two fixed points. In other words, the atoms change their positions (giving high RMSF) but do that within a limited space (therefore the density remains traceable).



a **b**

Figure 7. **a:** The nucleotides located deeply inside the virion (within 8 nm from the capsid centre, blue). The intervals, which had no experimental backbone, are highlighted in white. **b:** The nucleotides located near the capsid wall (beyond 8 nm from the capsid centre, orange). The intervals 2860–2915 and 3115–3195 are highlighted in green. Blue sphere encompasses the genome within 8 nm from the capsid centre.

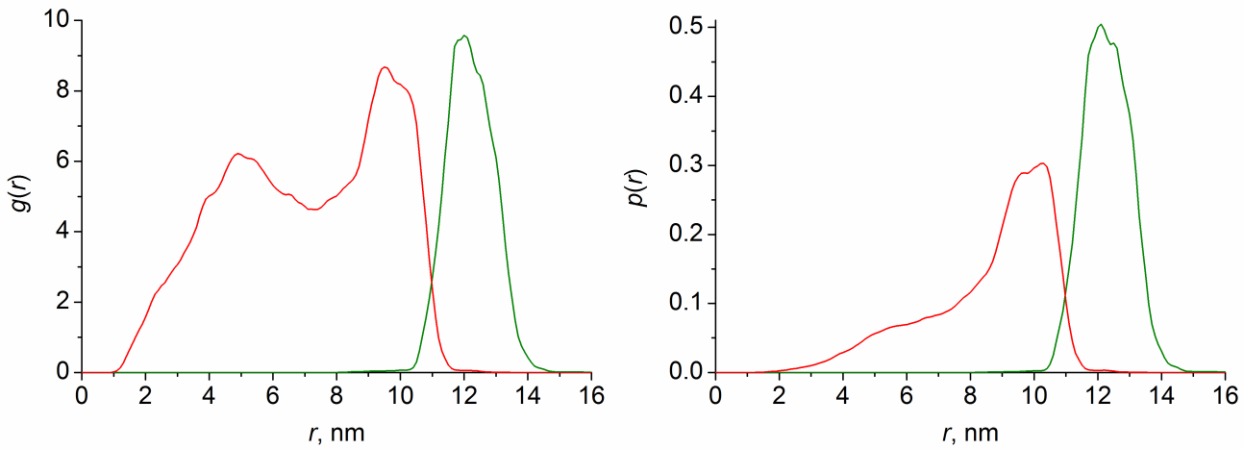
Overall, the calculations suggest that the likely reasons for the gaps in the cryo-EM measurements are the flexibility of the macromolecule and the higher mobility of some of its segments.

3.3. Structural characteristics

A fundamental structural characteristic of the virus particle is the distribution of its RNA inside the capsid, which can be useful for theoretical research. [25] It can conveniently be described by a radial distribution function (RDF) $g(r)$ with respect to the particle centre, see Fig. 8a. This quantity describes the RNA density at a given distance from the centre. The function has two local maxima separated by a minimum, which indicates that the genome is organised in two layers: one is located deeply in the capsid, while the second one is aligned near the capsid wall. The proportion between the layers is shown by the distribution function $p(r)$, see Fig. 8b, that is related to $g(r)$ via Eq. 1. Most of the genome is collected near the capsid wall at 8–10 nm from the centre, but a considerable fraction stays deeper, at 5–6 nm. Specifically, 31% of RNA is located within 8 nm from the centre, and 69% is beyond this distance (this is determined by integrating $p(r)$ from $r = 0$ to $r = 8$ nm):

$$p(r) = C r^2 g(r), \qquad (1)$$

where C is a normalisation factor that renders the integral of $p(r)$ from $r = 0$ to $r = \infty$ to be equal to unity.



a **b**

Figure 8. Radial distribution functions **(a)** and normalized distribution functions **(b)** of the genome (red) and the capsid (green) atoms with respect to the capsid centre computed over the production run.

The RNA–RNA hydrogen bonding is of a separate interest because it determines the base pairing and secondary structure of the macromolecule. We monitored the change of the secondary structure of the genome during the simulation. The RNApdbee web server [26–28] was used for the analysis, for most options the default values were used, except for including non-canonical base pairs like G–U into the calculations. Individual configurations were extracted from MD trajectory, and no time averaging was done.

Overall, 11.9% of nucleotides changed their base pairing during the 200 ns run. For the intervals 0 – 100 ns and 100 – 200 ns, these numbers are 10.1% and 9.6%, respectively, showing that many of these changes are repeated in the two intervals. For comparison, we considered intervals of length 1 ns, namely, 100 – 101 ns and 199 – 200 ns. For these intervals, the numbers of involved nucleotides are 9.3% and 7.8%, respectively. While these numbers serve only as examples (compare with Fig. 5b), it should still be noted that a large fraction of the observed changes in the base pairing in the 0 – 200 ns interval are already observed in the 199 – 200 ns interval. This indicates that most changes of the base pairing are consistent with thermal motion and not with large-scale refolding. A visual inspection of the secondary structure during the simulation indeed shows that most of these changes occurred within stem-loops when few individual base pairs broke or formed. The dot-bracket representation of the structures is presented in Supporting Information.

3.4. Ion distributions

Three kinds of mobile ions are present in the simulated solution, and their number and distribution determine the total charge of the virus particle and its electrokinetic mobility, which can be macroscopically observed. We monitored the change of the total charge of the virus particle with inner and adjacent ions; it was calculated as the charge of the RNA and the capsid plus the charge of the ions within 14 nm from the capsid centre, see Fig. 9a. The value fluctuates within a short range with no apparent trend, indicating the correctness of the initial ion placement and completeness of the equilibration. The final charge of the complete virion in our conditions was found to fluctuate around -153 $|e|$ indicating negative surface potential and ζ-potential, in accordance with the available experimental information [29].

More detailed information is provided by the radial distribution functions of the ions with respect to the capsid centre, see Fig. 9b. Within the capsid (at $r < 10$ nm), the RDFs of the cations have numerous local minima and maxima despite long time and ensemble averaging. The explanation is their binding to RNA, which has a complex spatial pattern. For Cl^- anions this effect is minor because they are not aligned to the RNA. Mg^{2+} cations are strongly accumulated at $r = 8$–10 nm following the distribution of RNA (Fig. 8), while Cl^- are expelled from this space. Na^+ are distributed rather evenly within the 5–10 nm range. $g(r)$ of Cl^- is shifted to the right compared to $g(Na^+)$ because Cl^- are attracted to the positively charged inner side and repulsed from the negatively charged outer surface. In contrast, Na^+ ions form a peak at 12.6 nm corresponding to the ions placed in the concave regions of the outer surface, impregnated into it, and collected near pores. The following larger peak at $r = 14$ nm corresponds to the formation of a Na^+ layer around the virus particle having a net negative charge. Mg^{2+} were originally placed inside the capsid only and did not leave it in a significant quantity, therefore the corresponding value of $g(r)$ vanishes beyond $r = 11.5$ nm.

A comparison with the distribution of the ions in the empty capsid indicates a very similar character of $g(r)$ for Na^+. However, Cl^- anions had a high peak at $r = 10$–11 nm, which is suppressed here because of the displacement of Cl^- by the negatively charged RNA.

Although being a very general characteristic, radial distribution of ions is used in the thermodynamic description of virus particles and in the process of their assembly [30,31].

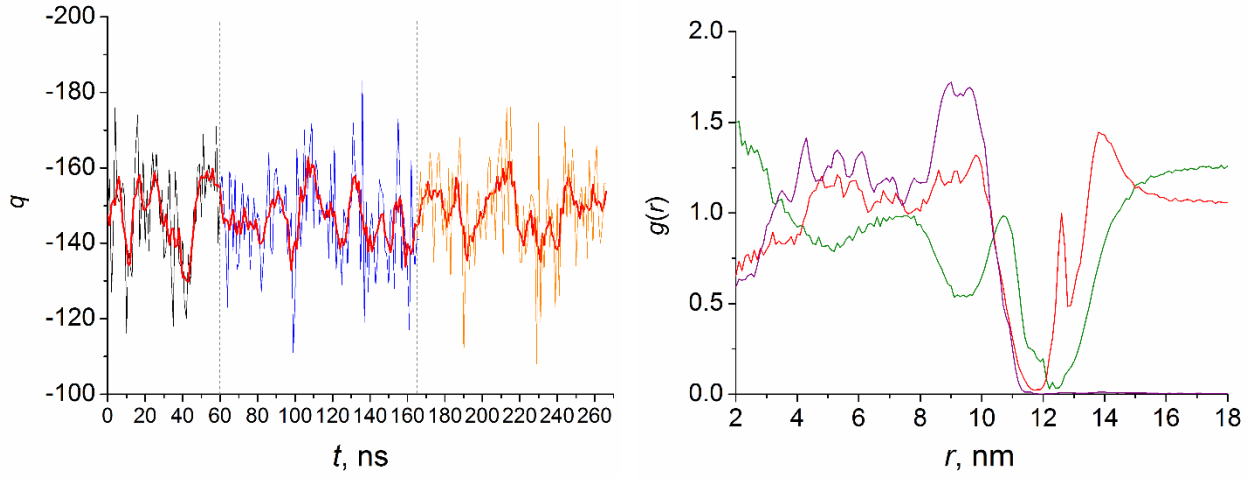


Figure 9. **a:** Time change of the total charge of the virus particle. The black curve is NVT equilibration, the blue curve is NPT equilibration, the orange curve is production run, the red curve is adjacent average over 5 ns. Dashed lines indicate the boundaries of the runs. **b:** Radial distribution functions of Na^+ (red), Mg^{2+} (purple), and Cl^- (green) ions with respect to the capsid centre computed over the production run. The curve for Mg^{2+} is scaled down 5 times for visual purpose.

3.5. Solution transport

The capsid contains 30 pores located at the junctions of three or five capsid proteins. Additionally, there is a hole made by two missing capsid proteins, which are replaced by the maturation protein crossing the capsid wall. There is some spacing between the capsid and MP, which serves as a kind of extra pore. The diameter of the pores depends on its location (the ones located at the junctions of five capsomers are larger); overall it varies from pore to pore within the range of 0.8 – 1.4 nm with the most frequent diameter close to 1 nm, Fig. 10a. The measurements were done via visual examination in VMD because of the difficulty in rigorous quantification of this characteristic.

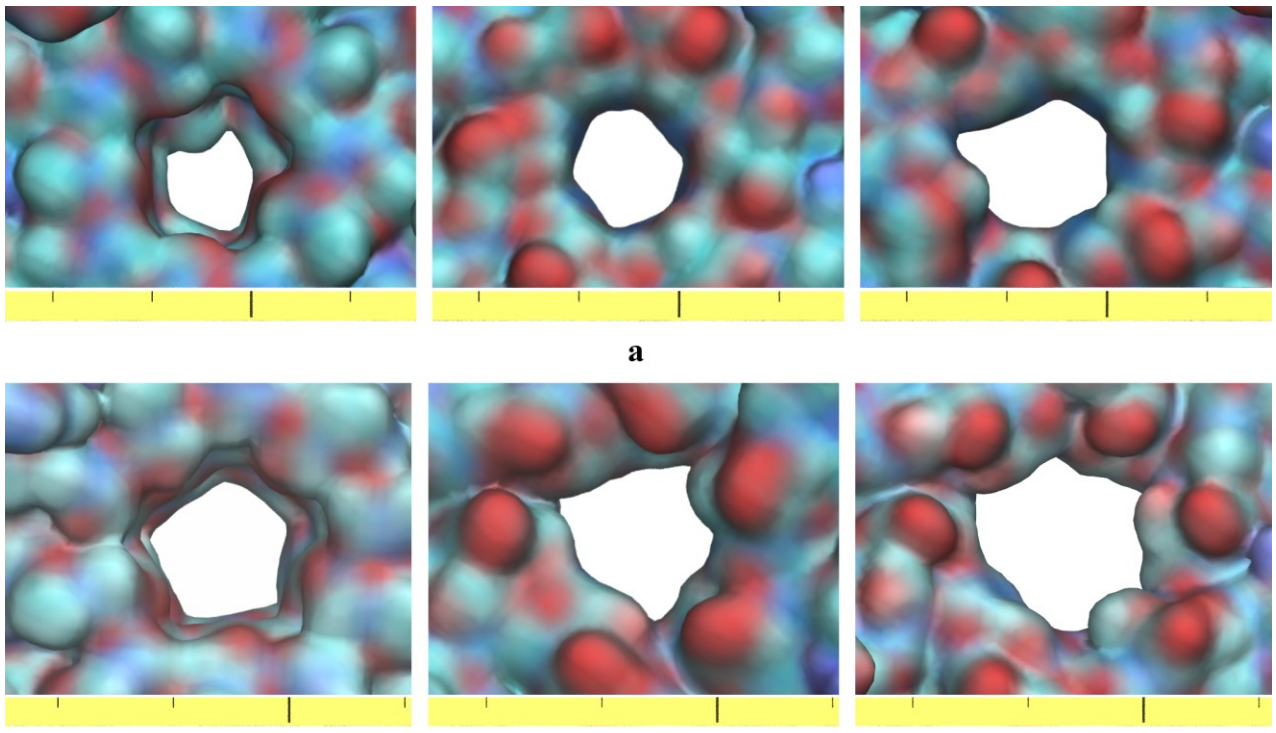
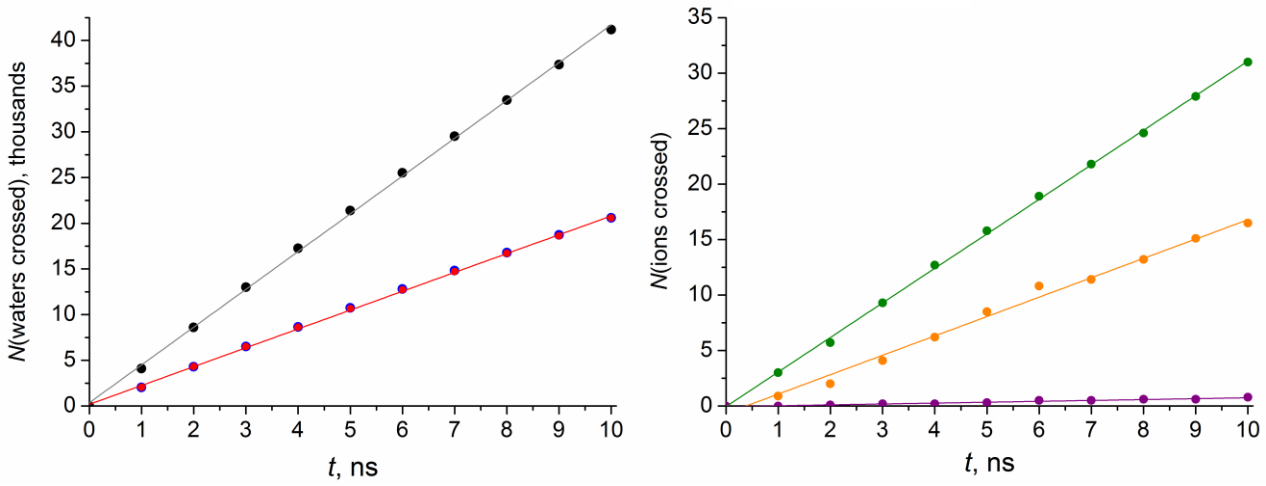


Figure 10. **a:** Pores in the virion (AMBER14sb force field); **b:** Pores in the empty capsid (AMBER03 force field). Examples of relatively small, intermediate, and relatively large pores are shown on the left, in the middle, and on the right, respectively. 1 division of the scale is 1 nm.

As a result, the capsid serves as a semipermeable membrane, allowing limited transfer of species to and from its interior. The MD simulation allows investigating this feature quantitatively. We computed the flow rate of water and ions across the capsid. The algorithm used by us previously for the empty MS2 capsid was followed, which itself grounds on the algorithm proposed by Andoh *et al.* [32]. The outline is as follows: a configuration at time t_0 was picked, and the ions and water O atoms located inside and outside the capsid were enumerated. Because of almost spherical shape of the capsid, the sorting was based on the distance of the atom to the capsid centre: it was viewed as located inside or outside if the distance was shorter than 11.7 nm or longer than 12.7 nm, respectively. These borders equal to the ones we used for the empty MS2 and were chosen to match the average radii of the inner and outer surfaces, if visualised in VMD. The 1 nm thick buffer is needed to exclude the atoms stuck in the pores or impregnated in the wall. The exact value of the thickness did not influence significantly the computed flows.

Then, the configurations at $t_1 = t_0 + \Delta t$ were picked, and the lists of the atoms inside and outside were composed. The atoms present in the former list and absent in the latter one were considered as having left during Δt, and those absent in the former but present in the latter were considered as having entered during Δt. The sum of left and entered atoms gave the total amount of atoms crossed the boundary. Δt varied from 1 to 10 ns and t_0 was chosen equal 0 ns, 10 ns, ... 90 ns. The results calculated for different t_0 were then averaged, and they are presented in Fig. 11.



a **b**

Figure 11. **a**: The average number of water molecules, entering the capsid during the given time interval Δt (red circles), leaving it (blue circles), or crossing the boundary in both directions (black circles). Red and blue circles almost coincide. **b**: The average number of Mg^{2+} (purple), Na^+ (orange), and Cl^- (green) ions crossing the capsid boundary in both directions during the given time. Linear regression curves have been added to all data.

On average, 4140±40 water molecules crossed the capsid border each nanosecond. The ingoing and outgoing flows were almost equal indicating the equilibration of the model. The flows of Na^+ and Cl^- were 1.7±0.1 and 3.1±0.1, while the transport of Mg^{2+} was negligible in the simulation (it can be estimated as 0.08±0.01 ions per ns). Compared to the empty capsid described with the same force field, it is ca. 20% reduction for water and ca. 50% reduction for Cl^-, Table 1. The flow of Na^+ cations is equally small in both cases. Importantly, the diameter of pores did not change significantly upon the introduction of the genome. Hence, we suggest that the genome sterically blocks some fraction of pores reducing the flow of solution. Overall, the transport properties of the capsid were moderately reduced by the introduction of the genome.

It is appropriate here to compare the transport characteristics of the empty capsid computed in two different force fields, AMBER14sb and AMBER03 (as in Ref. [9]). They are compared in Table 1. The model made using the new force field shows considerably weaker flow of water (by ca. 15%). We identify the main factor to be the size of the pores that is smaller in the AMBER14sb model (Fig. 10b). For the ions the effect of the new forcefield was controversial: the flow of Na^+ reduced, as well; while the flow of Cl^- somewhat increased in magnitude. We attribute this to the decreased pore size and also to the effect of CUFIX correction, which we applied over AMBER14sb force field. There, the Na^+–H_2O interaction was made stronger, while the Na^+–Cl^- one was weakened. In total this may reduce ionic pairing between Na^+ and Cl^- and thus boost Cl^- mobility.

Table 1. The transport characteristics of the empty and genome-loaded MS2 capsids.

characteristics	genome-loaded capsid, AMBER14sd	empty capsid, AMBER14sd	empty capsid, AMBER30 [9]
water flow, molecules per ns	4140	5040	5800
Na^+ flow, ions per ns	1.7	1.6	4.5
Cl^- flow, ions per ns	3.1	6.4	4.9

Conclusions

We have built a complete atomistic model of an entire virus particle, the bacteriophage MS2, consisting of both native genome and native capsid including gaps-free maturation protein. The structure of the genome at room temperature and in physiological solution was found stable and it remained largely the same as the initial structure reconstructed from the cryo-EM measurements.

We have analysed structural and dynamic properties of the virus particle and compared them with those of a genome free MS2 virus like particle. The distribution of ions inside and outside the capsid revealed a pronounced layer of chloride ions at the inner surface, while the sodium ions accumulated at the outside surface of the capsid. Finally, the pores in the capsid wall sustain substantial transport of water and ions in both directions.

Through a detailed analysis of the simulations, we were able to explain the gaps in the experimental cryo-EM data using the structural and spatial information of the simulated genome. Furthermore, we observe that thermal motion changes the secondary structure of the genome constantly via the formation of bonds within stem-loops, but without leading to large scale re-folding of the RNA secondary structure.

Acknowledgements

V. F. thanks the Ministry of Education and Science of Ukraine for financial support in the frame of the project #0120U101064. V. F. and D. N. acknowledge the use of HPC Midlands supercomputer funded by EPSRC, grant number EP/P020232/1; the access to HPC Call Spring 2021, EPSRC Tier-2 Cirrus Service; the access to Sulis Tier 2 HPC platform hosted by the Scientific Computing Research Technology Platform at the University of Warwick. Sulis is funded by EPSRC Grant EP/T022108/1 and the HPC Midlands+ consortium. The collaboration was supported by the program H2020-MSCA-RISE-2018, project AMR-TB, Grant ID: 823922. We acknowledge support from the EPSRC grant EP/M02735X/1 (AMR4AMR).

Associated content

Supporting Information available. Secondary structure of MS2 genome during MD simulation in dot-bracket notation at 0, 165, 166, 264, and 265 ns of the course of MD simulations.

References

(1) Tarasova, E.; Nerukh, D. All-Atom Molecular Dynamics Simulations of Whole Viruses. *J. Phys. Chem. Lett.* **2018**, *9*, 5805–5809.
(2) Machado, M. R.; Pantano, S. Fighting viruses with computers, right now. *Current Opinion in Virology* **2021**, *48*, 91–99.
(3) Larson, S.; McPherson, A. Satellite tobacco mosaic virus RNA: structure and implications for assembly. *Curr. Opin. Struct. Biol.* **2001**, *11*, 59–65.
(4) Kappel, K.; Liu, S.; Larsen, K. P.; Skiniotis, G.; Puglisi, E. V.; Puglisi, J. D.; Hong Zhou, Z.; Zhao, R.; Das, R. De novo computational RNA modeling into cryo-EM maps of large ribonucleoprotein complexes. *Nat. Methods* **2018**, *15*(11), 947–954.
(5) Strauss, J. J.; Sinsheimer, R. L. Purification and properties of bacteriophage MS2 and of its fibonucleic acid. *J. Mol. Biol.* **1963**, *7*, 43–54.
(6) Hu, B.; Khara, P.; Christie, P. Structural bases for F Plasmid conjugation and F pilus biogenesis in Escherichia coli. *Proc. Natl. Acad. Sci. (PNAS)* **2019**, *116*, 14222–14227.
(7) Harb, L.; Chamakura, K.; Khara, P.; Christie, P.; Young, R.; Zeng, L. ssRNA phage penetration triggers detachment of the F-pilus. *Proc. Natl. Acad. Sci. (PNAS)* **2020**, *117*, 25751–25758.
(8) Wagner, A.; Weise, L.; Mutschler, H. InVitro characterisation of the MS2 RNA polymerase complex reveals host factors that modulate emesviral replicase activity. *Comm. Biol.* **2022**, *5*, 264–275.

(9) Farafonov, V. S.; Nerukh, D. MS2 bacteriophage capsid studied using all-atom molecular dynamics. *Int. Focus* **2019**, *9*(3), 20180081.
(10) Farafonov, V. S.; Stich, M.; Nerukh, D. Reconstruction and validation of entire virus model with complete genome from mixed resolution cryo-EM density. *Faraday Discuss* **2022**, *240*, 152–167.
(11) Farafonov, V. S.; Vodolazkaya, N.; Nikolskaya, M.; Laguta, A.; Balklava, Z.; Stich, M.; Mchedlov-Petrossyan, N.; Nerukh, D. All-atom structure and dynamics of entire virus with genome in solution at physiological conditions. *Faraday joint interest group conference.* online, **2021**.
(12) Farafonov, V.; Bakumenko, M. Y.; Bardik, V.; Pozhilenkov, A.; Kalachev, A.; Vaysfeld, N.; Stich, M.; Nerukh, D. All-atom MD of entire virus with genome in solution at physiological conditions. *WATOC 2020, 12th Triennial Congress of the World Association of Theoretical and Computational Chemists.* Vancouver, Canada, **2022**.
(13) Available at https://avogadro.cc. Accessed by 02 Aug 2023.
(14) Available at https://manual.gromacs.org/documentation/2022-current/reference-manual/special/pulling.html. Accessed by 02 Aug 2023.
(15) Tarasova, E.; Farafonov, V.; Khayat, R.; Okimoto, N.; Komatsu, T.; Taiji, M.; Nerukh, D. All-atom molecular dynamics simulations of entire virus capsid reveal the role of ion distribution in capsid's stability. *J. Phys. Chem. Lett.* **2017**, *8*(4), 779–784.
(16) Fischer, N. M.; Polêto, M. D.; Steuer, J.; van der Spoel, D. Influence of Na^+ and Mg^{2+} ions on RNA structures studied with molecular dynamics simulations. *Nucleic Acids Res.* **2018**, *46*(10), 4872–4882.
(17) Maier, J. A.; Martinez, C. K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*(8), 3696–3713.
(18) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham III, T. E.; Jurečka, P. Refinement of the Cornell et Al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7*(9), 2886–2902.
(19) Sponer, J.; Bussi, G.; Krepl, M.; Banas, M.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurečka, P.; Walter, N. G.; Otyeptka, M. RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chem. Rev.* **2018**, *118*, 4177−4338.
(20) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.
(21) Yoo, J.; Aksimentiev, A. Improved Parametrization of Li+, Na+, K+, and Mg2+ Ions for All-Atom Molecular Dynamics Simulations of Nucleic Acid Systems. *J. Phys. Chem. Lett.* **2012**, *3*, 45–50.
(22) Yoo, J.; Aksimentiev, A. Refined Parameterization of Non-Bonded Interactions Improves Conformational Sampling and Kinetics of Protein Folding Simulations. *J. Phys. Chem. Lett.* **2016**, *7*(19), 3812–3818.
(23) Yoo, J.; Aksimentiev, A. New tricks for old dogs: Improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *Phys. Chem. Chem. Phys.* **2018**, *20*(13), 8432–8449.
(24) Available at https://bionano.physics.illinois.edu/CUFIX. Accessed by 02 Aug 2023.
(25) Belyi, V. A.; Muthukumar, M. Electrostatic origin of the genome packing in viruses. *Proc. Natl. Acad. Sci.* **2006**, *103*, 17174–17178.
(26) Antczak, M.; Popenda, M.; Zok, T.; Zurkowski, M.; Adamiak, R. W. New algorithms to represent complex pseudoknotted RNA structures in dot-bracket notation. *Bioinformatics* **2017**, *34*(8), 1304–1312.
(27) Zok, T.; Antczak, M.; Zurkowski, M.; Popenda, M.; Blazewicz, J.; Adamiak, R. W.; Szachniuk, M. RNApdbee 2.0: multifunctional tool for RNA structure annotation. *Nucleic Acids Res.* **2018**, *46*(W1), W30–W35.
(28) Antczak, M.; Zok, T.; Popenda, M.; Lukasiak, P.; Adamiak, R. W.; Blazewicz, J.; Szachniuk, M. RNApdbee – a webserver to derive secondary structures from pdb files of knotted and unknotted RNAs. *Nucleic Acids Res.* **2014**, *42*(W1), W368–W372.
(29) Vodolazkaya, N.; Nikolskaya, M.; Laguta, A.; Farafonov, V.; Balklava, Z.; Stich, M.; Mchedlov-Petrossyan, N.; Nerukh, D. Estimation of nanoparticle's surface electrostatic potential in solution using acid-base molecular probes. III: Experimental hydrophobicity/hydrophilicity and charge distribution of MS2 virus surface. *J. Phys. Chem. B* **2022**, *126*(41), 8166−8176.
(30) Ting, C. L.; Wu, J.; Wang, Z.-G. Thermodynamic basis for the genome to capsid charge relationship in viral encapsidation. *Proc. Natl. Acad. Sci.* **2011**, *108*, 16986–16991.

(31) Bozic, A. L.; Siber, A.; Podgornik, R. How simple can a model of an empty viral capsid be? Charge distributions in viral capsids. *J. Biol. Phys.* **2012**, *38*, 657–671.
(32) Andoh, Y.; Yoshii, N.; Yamada, A.; Fujimoto, K.; Kojima, H.; Mizutani, K.; Nakagawa, A.; Nomoto, A.; Okazaki, S. All-atom molecular dynamics calculation study of entire poliovirus empty capsids in solution. *J. Chem. Phys.* **2014**, *141*, 165101.

For Table of Contents Only