MS2 bacteriophage capsid studied using all-atom Molecular Dynamics

Vladimir S. Farafonov[1], Dmitry Nerukh[2 *]

[1] Department of Physical Chemistry, V. N. Karazin Kharkiv National University, 61022, Kharkiv, Ukraine, [2] Systems Analytics Research Institute, Department of Mathematics, Aston University, Birmingham, B4 7ET, UK
*corresponding author: nerukhd@aston.ac.uk

ABSTRACT

The all-atom model of an MS2 bacteriophage particle without its genome (the capsid) was built using high resolution cryo-EM measurements for initial conformation. The structural characteristics of the capsid and the dynamics of the surrounding solution were examined using Molecular Dynamics simulation. The model demonstrates the overall preservation of the cryo-EM structure of the capsid at physiological conditions (room temperature and ions composition). The formation of a dense anion layer near the inner surface and a diffuse cation layer near the outer surface of the capsid was detected. The flow of water molecules and ions across the capsid through its pores were quantified, which was considerable for water and substantial for ions.

Keywords: MS2 virus capsid, ions distribution, flow of solution.

1. Introduction

State-of-the-art computers can model very large biomolecular systems of millions of atoms in size (organelles, cellular membranes, entire viruses, etc) at atomistic resolution [1-3]. The required initial molecular structure of the biomolecule itself, the 3D coordinates for all atoms in the system, is derived from high resolution experimental techniques such as X-ray crystallography and cryo-electron microscopy (cryo-EM). All-atom Molecular Dynamics (MD) computer models can provide information about, and unique insights into, complex biomolecular systems, which are impossible to obtain by any other means. For example, the reconstruction of the parts of the molecular system, such as highly flexible regions, that cannot be revealed experimentally, provides an understanding of dynamics at physiological temperature. Importantly, while until recently X-ray crystallography was the main source of molecular structures, nowadays cryo-EM measurements provide information of comparable resolution but with a critical advantage of being able to measure asymmetric structures.

Simulation of viruses is most realistic because, in contrast to other large biomolecular systems, viruses are self-contained biological units, which exist in isolation from the rest of the organism (although they cannot reproduce in isolation). Computer models of cellular organelles, for example, necessarily consider only part of the system that interacts with the rest of the cell via complicated, poorly understood mechanisms making comparison with the experimental results difficult. For viruses, interaction with aqueous solution is the only external force that defines the structure and dynamics, and modelling of water is well developed in MD. Despite the substantial computing resources needed for such simulation, the number of viruses simulated at all-atom resolution is growing (see the reviews on all known simulations up to date [4,5]).

We have recently built an MD model of PCV2 virus capsid, which is the smallest known non-satellite virus (it does not need another virus for reproduction) [6-8]. Here we report on MD modelling of a well-known bacteriophage that infects *E. coli*, the (+)ssRNA phage MS2. This virus is a perfect candidate for such all-atom simulations as it is small (it has a short genome of 3569 nt), its icosahedral capsid is only marginally larger than PCV2 (27 nm vs 23 nm) but it is much more extensively investigated using various molecular biology and biophysical methods [9-11].

The molecular structure of the MS2 capsid has been measured recently with atomistic resolution providing not only the structure of the symmetric parts of the capsid, but also several proteins that are different from the rest of the capsid, making it asymmetric [12]. The most recent publication even provides parts of the genome resolved at atomistic level and most of the genome at the resolution allowing to fit its backbone [13]. We are, thus, at a unique position of being able to build a very realistic MD model of the virus capsid including its asymmetric parts that are known to play an important biological role in infecting bacteria.

The MS2 capsid mostly consists of multiple copies of one protein 129 amino acid residues long. The protein adopts three different conformations, denoted A, B, and C, which form two kinds of dimers: asymmetric AB and symmetric CC. The capsid is composed of 59 AB and 30 CC dimers. Finally, there is a 393 residues long maturation protein embedded in the capsid wall [14]. The capsid forms an impenetrable wall with pores that allow water and ion transport.

We have used this information to develop the MD model of the capsid and analyse a number of its structural and dynamic properties. We have not included the genome in the model, however this work is in progress and the results will be the subject of our subsequent publications.

2. Simulation details

2.1 Assembling the capsid model

Our capsid model is based on the cryo-EM structure 5TC1 from the Protein Data Bank [15]. It contains i) three coat protein monomers, forming a triangle-shaped block; ii) the maturation protein, of a sequence different from the rest of the capsid proteins; and iii) 5 monomers located near the maturation protein, whose structure is deformed by the maturation protein. Among these monomers, there is a gap in structure information in one of them (denoted chain G in the PDB file), that is the coordinates of the residues 68–77 are missing, probably because of their flexible or disordered structure. Furthermore, there are also 4 gaps in the maturation protein (63 residues in total).

The details of the assembling procedure are presented in Supplementary Information. GROMACS 5.1 simulation package was employed for all the tasks [16]. The final simulation cell is depicted in Fig. 1.

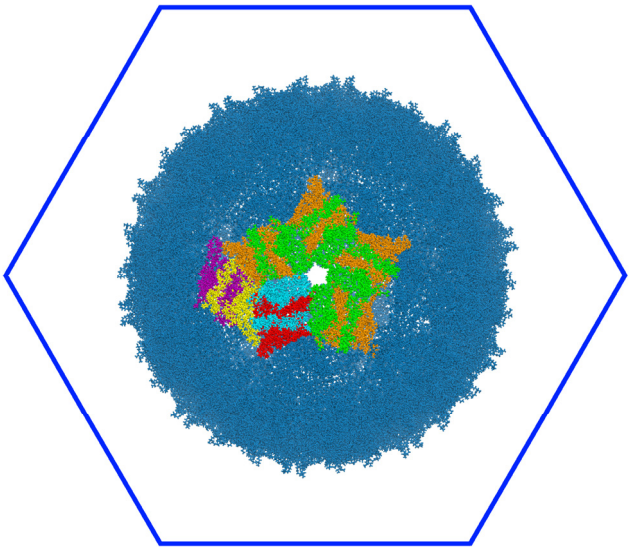

Figure 1. The capsid model in a rhombic dodecahedron shaped simulation box. Several coat proteins are coloured highlighting the pentamer enclosing the pore (green, orange, blue and red molecules) and the trimer block (green, orange, blue, red, purple, and yellow molecules). Ions and water molecules are not shown.

2.2 Simulation protocol

The simulations were preformed using AMBER03 force field and TIP3P water model. The total number of atoms in the system was 3528811 including 1057725 water molecules.

First, the energy minimisation of the assembled system was done using the steepest descent algorithm until the largest residual force became less than 5000 kJ nm^{-1}. Then, 5 ns equilibration run was carried out at constant volume. The heavy atoms of the capsid were restrained to their initial positions except for the reconstructed missing sequence, which was allowed to adopt the optimal structure. The initial temperature was set to 100 K and during the first 200 ps it was gradually raised to 298 K; this value was kept for the rest of the run (simulated annealing options of GROMACS were used). The average pressure in the cell was found to be 120 bar, which is close to the standard pressure. Actually, according to the compressibility of water at 300 K ($4.5 \cdot 10^{-5}$ bar^{-1} [16]), this extra pressure corresponds to cell length change 0.07 nm out of 37 nm that is negligible. Thus, we choose not to employ barostat for the production run in order to avoid introducing unwanted volume fluctuations to the system.

After equilibration a productive run of 50 ns long was carried out without restraints. The volume of the cell was kept constant. The average pressure during the run was –85 bar.

In all runs 3D periodic boundary conditions were imposed, the temperature was controlled by velocity-rescale thermostat with time constant of 1 ps, the time step was 2 fs, all bonds were constrained with LINCS algorithm, the electrostatic interactions were computed with the PME method (cut-off for direct sum was at 1 nm), while van der Waals interactions were cut-off at 1 nm.

3. Results and discussion

3.1 Structure

The basic indicator of the stability of the structure is the time evolution of its root mean square deviation (RMSD) from the initial configuration. The graphs for the whole capsid and its backbone only are shown in Fig. 2A. Both RMSD's increased for 20 ns and then reached a plateau at a small value (~0.27 nm) indicating no considerable displacements occurred. The same

is demonstrated by the time evolution of the capsid's radius of gyration (Fig. 2B): it essentially stops growing after 20 ns showing ~0.15 nm increase at the end of the simulation compared to the cryo-EM structure.

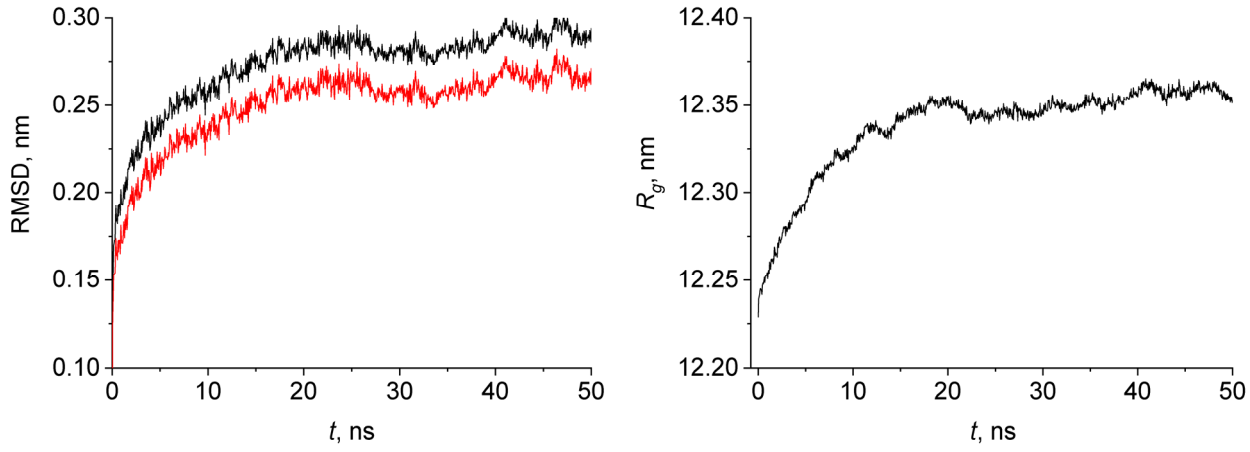

Figure 2. A: RMSD of the whole capsid (black) and its backbone (red) from the initial structure. **B**: The radius of gyration of the capsid.

It is interesting to inspect the average displacement of individual amino acid residues from their initial positions, Fig. 3. All residues have similar low average displacement (in terms of RMSD) with notable exceptions being residues 13-15 with the values of ~0.8 nm. These residues are located on the outer surface far from the pores and they are exposed to the outside solution, which makes them relatively more mobile (Fig. 4).

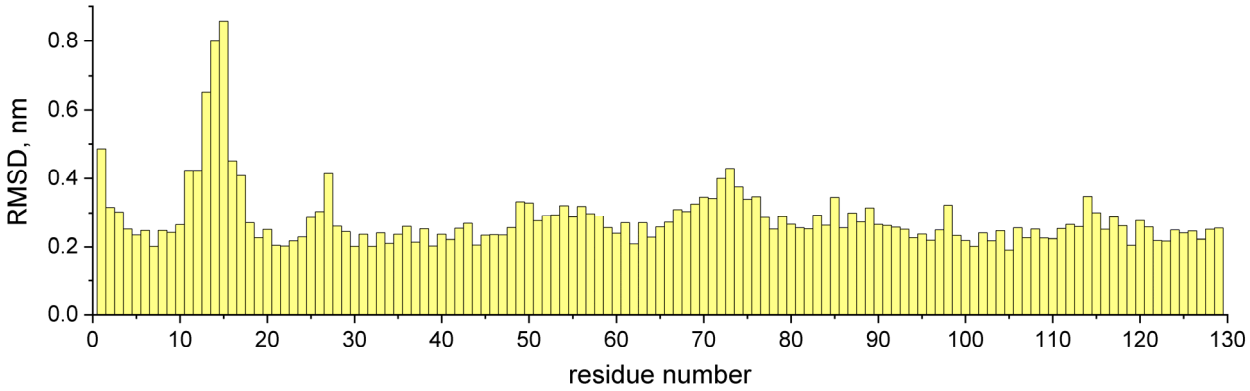

Figure 3. RMSD of individual residues averaged across 89 coat protein dimers.

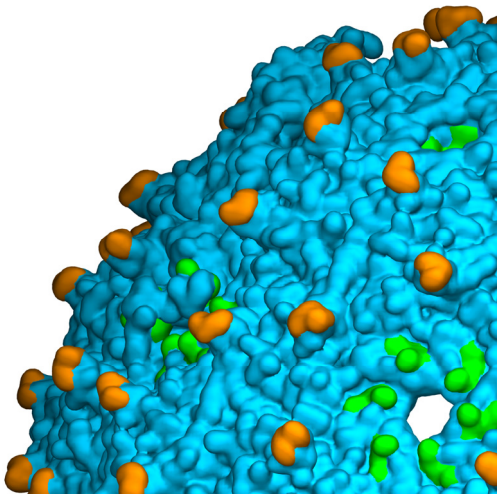

Figure 4. The initial location of residues 13–15 having the largest displacement from the cryo-EM structure (orange) and 75–78 forming the FG loop (green, see text).

From the biological point of view, the secondary structure (SS) of the coat protein is important. MD simulation allows verifying whether the SS at the cryo-EM conditions is preserved at room temperature. For this, the DSSP 2.0.4 software [16] was used.

We computed the 'average' secondary structures of A-, B-, and C-kind coat protein monomers over all copies in i) the cryo-EM configuration, ii) the configuration after the restrained run (that is the initial conformation for the productive one), iii) a series of intermediate configurations at 5–45 ns, and iv) the final configuration, Fig. 5. Specifically, for each 1-129 amino acid residue the SS that is the most populated in monomers was determined; and the sequence of 129 single-residue SS'es formed the 'average' SS. In total, in the final configuration 16% (A-kind monomers) and 13% (B- and C-kind monomers) of amino acid residues changed their average SS after 50 ns of simulation compared to the cryo-EM conformation.

A crucial region of the coat protein is the so-called FG loop (residues 75-78, Fig. 5) that plays a key role in the self-assembly of the capsid [17]. Its secondary structure in A- and C-kind monomers is identical and different from that in B-kind ones. The inspection shows that the SS of this interval in B-kind monomers stayed unchanged during the whole simulation, while in both A- and C-kind ones, an reversible modification at residues 74–75 occurs. However, the adjacent interval 79–81 or 79–82 has stably restructured in A- and B-kind monomers, respectively.

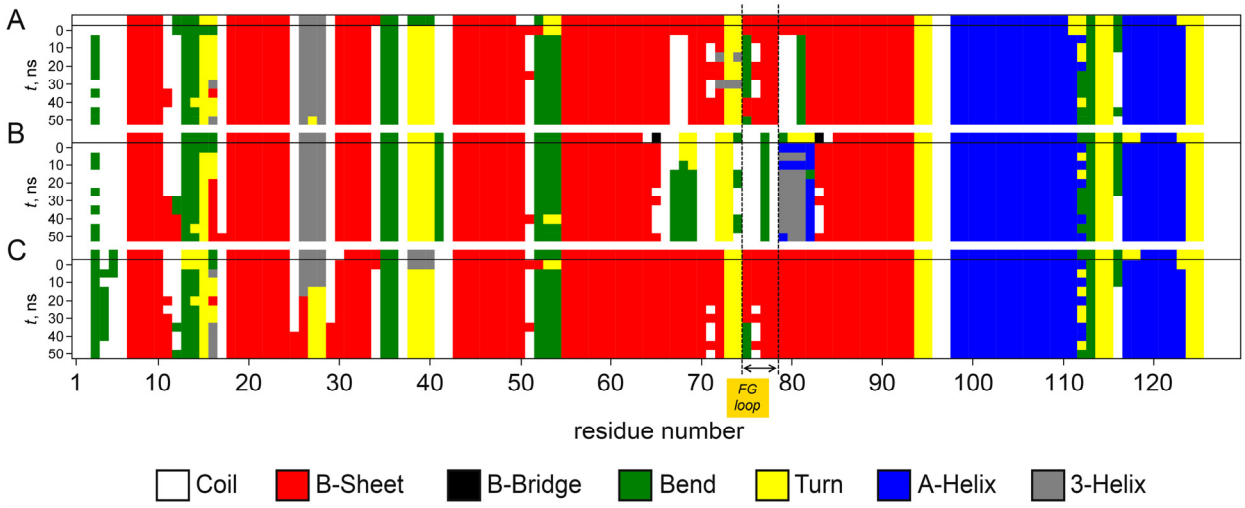

Figure 5. Capsid averaged secondary structures of A-, B-, and C-kind coat protein monomers at different simulation times. For each monomer kind, the first row that is underlined corresponds to the cryo-EM configuration. The FG loop is indicated.

3.2 Ions distribution

The MS2 capsid is a highly charged particle. Therefore, the formation of an electric double layer in salt solution is expected. We computed the distribution of the ions around the capsid in terms of the radial distributions functions (RDF's) with respect to the capsid centre of mass (COM), Fig. 6. The RDF's on the intervals 10–30 ns and 30–50 ns are almost identical proving that the equilibrium distribution of the ions was reached.

Apparently, Cl^- ions form a distinct layer near the inner surface with the average density 2.1 times higher than in the inner solution. Surprisingly, Na^+ ions have very low concentration near the negative outer surface (~1.3 times). This is different compared to PCV2, where twofold increase is observed [8]. A small peak at 12.6 nm is formed by the ions located in pores impregnated in the wall or located near the wall at concave regions, while the larger peak at 14.2 nm is made of ions near the convex regions and ions in the solution several water layers away from the surface. Thus, the whole virus-like particle (VLP) would behave as negatively charged, that corresponds to the experimental data showing negative zeta-potential values of MS2 VLP's [17-19].

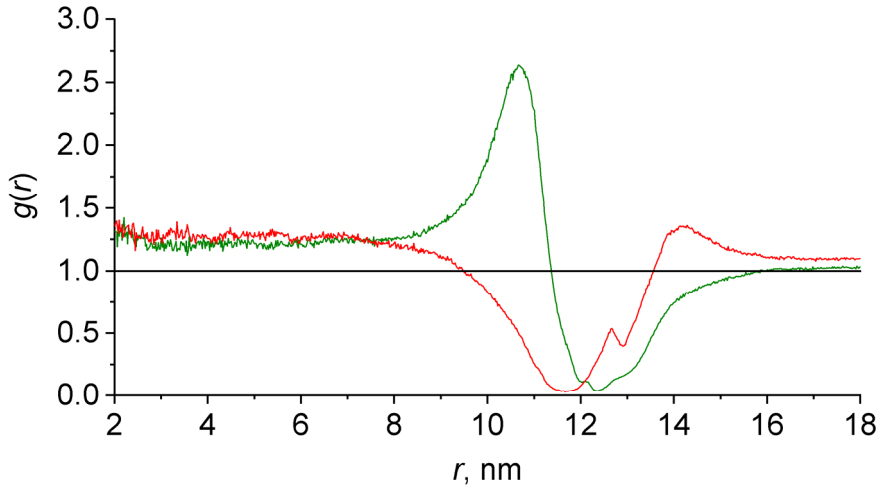

Figure 6. Radial distribution functions of Na^+ (red) and Cl^- ions (green) with respect to the capsid COM computed over the production run.

3.3 Water and ion flows through the capsid

The ability to allow or block the flow of water molecules and ions is an important property of a biomolecular wall. Therefore, it is interesting to examine the capsid wall in this respect. MD provides direct way to quantitatively estimate the movement of solution across the wall.

The algorithm similar to our previous work [7], which is grounded on the work of Andoh et al. [1], was used. Because the MS2 capsid shape is much closer to a spherical layer than that of PCV2 capsid, the boundary between the interior and the outside solution can be defined simply as the certain distance from the COM.

An atom was considered as located in the interior if its distance to the capsid COM was less than 11.7 nm, and it was located in the outer solution if the distance was more than 12.7 nm. These numbers were deduced from visual examination of the capsid in VMD to match the average radii of inner and outer surfaces. We did not make the radii to be functions of time, because the capsid size stayed almost constant after 10 ns (Fig. 2b, R_g has changed by <0.03 nm). The 1 nm buffer zone was introduced to exclude the molecules staying in the pores and to count only the molecules actually transferred through the capsid wall. As it will be shown below, the magnitude of the computed flows does not change significantly with the buffer zone thickness. Using this criterion the lists of indices of water molecules and ions located inside were formed. The particles inside (outside) the capsid at a time moment and outside (inside) at another time moment were considered to have left (entered) the capsid; and the sum of the left and entered particles gave the net flow across the wall during the time interval between the considered moments.

The calculation of transported particles was performed independently on five 10 ns intervals of the trajectory (0–10 ns, 10–20 ns, etc.). The average and standard deviation were computed on the obtained numbers; the results for the 0–10 ns interval appeared somewhat outstanding (likely because of residual equilibration of the capsid size, Fig. 2b), thus, they were excluded from the sampling. The graphs for the last interval (40–50 ns) are shown in Fig. 7; graphs for other intervals are very similar.

There is an intense flow of water molecules across the wall, reaching (5800±85) molecules per ns. The magnitude of the value does not depend significantly on the buffer thickness taken: the flow equals (5970±80) or (5630±85) molecules per ns if the buffer zone is placed at 12–12.5 nm or 11.5–13 nm, respectively. In the first 15 ns, ~6800 water molecules has left the capsid interior, while after that time the ingoing and outgoing flows became mutually compensating.

The total flow of ions is considerable, too, reaching (4.5±0.5) Na^+ and (4.9±0.4) Cl^- ions per ns. The values do not change considerably when buffer thickness is increased or decreased. For Cl^- and, to less extent, Na^+, the outgoing flow is somewhat stronger than the ingoing one: after 50 ns, the content of the ions inside changed by –37 for Cl^- and –12 for Na^+.

The situation is completely opposite to the PCV2 capsid, where water transport was 70 times lower, and ions transport was negligible [7]. This is explained by the larger pore diameter in the MS2 capsid, reaching approximately 15 nm in size.

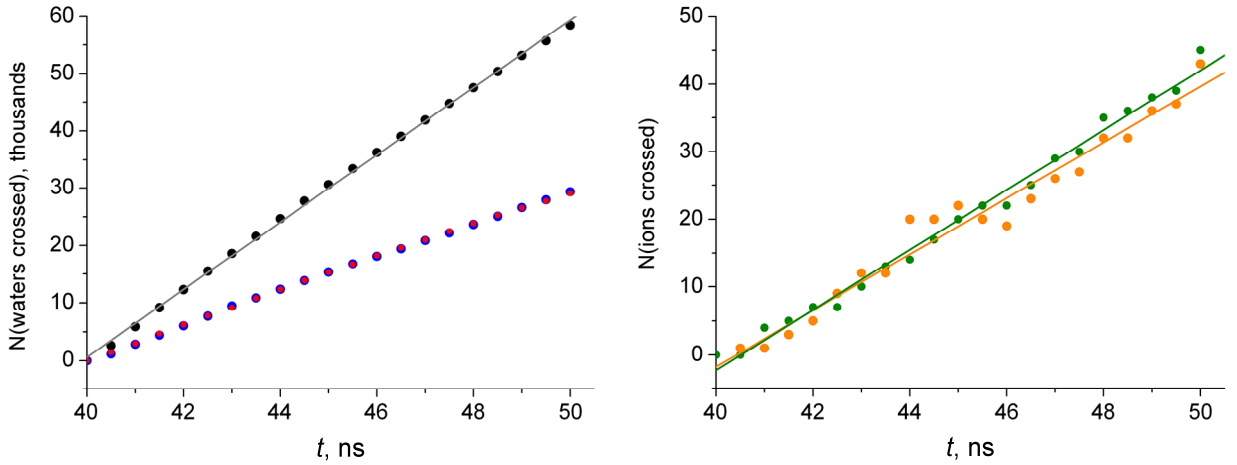

Figure 7. A: The number of water molecules that entered (red) or left (blue) the capsid interior compared to the structure at 40 ns, the sum of entered and left (black) and its linear fit (grey). **B:** The number of Na^+ (orange) or Cl^- (green) ions crossed the capsid wall in both directions compared to the initial structure, and its linear fit.

4. Conclusions

We have built an MD model of MS2 capsid, including its asymmetric parts (the maturation protein and several deformed proteins around) and one subsequence missing in the experimental data. After initial adjustment, the atomistic structure of the capsid at room temperature and in physiological solution was found stable and it remained largely the same as the measured cryo-EM structure with a small exception of three residues located on the outer surface.

We have analysed several structural and dynamics properties of the capsid that is impossible to deduce experimentally. The distribution of ions inside and outside the capsid revealed a pronounced layer of chloride ions at the inner surface, while the sodium ions formed a surprisingly diffuse layer at the outside surface of the capsid. Finally, the pores in the capsid wall sustain substantial transport of water and ions in both directions.

Data accessibility. This article has no additional data.

Authors' contributions. DN contributed to formulating the idea of the research and writing the text of the manuscript. VF performed simulations and analysis and contributed to writing the text of the manuscript.

Competing interests. We declare we have no competing interests.

Acknowledgements. The authors thank Dr Michael Stich for productive discussions.

Funding. V. F. acknowledges the support of the Ministry of Education and Science of Ukraine, grant number 0117U004966. D.N. acknowledges the support of JSPS through BRIDGE Fellowship, grant number BR170303. The authors acknowledge the use of HPC midlands supercomputer funded by EPSRC, grant number EP/P020232/1.

References

1. Andoh Y, Yoshii N, Yamada A, Fujimoto K, Kojima H, Mizutani K, Nakagawa A, Nomoto A, Okazaki S. 2014 All-atom molecular dynamics calculation study of entire poliovirus empty capsids in solution, *J. Chem. Phys.* **141,** 165101. (doi:10.1063/1.4897557)
2. Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, Ning J, Ahn J, Gronenborn AM, Schulten K, Aiken C, Zhang P. 2013 Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*, **497**, 643–6. (doi:10.1038/nature12162)
3. Perilla JR, Schulten K. 2017 Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations. *Nat. Commun*., **8**, 15959. (doi:10.1038/ncomms15959)
4. Hadden JA, Perilla JR. 2018 All-atom virus simulations. *Current Opinion in Virology*, **30**, 82–91. (doi: 10.1016/j.coviro.2018.08.007)
5. Tarasova E, Nerukh D. 2018 All-Atom Molecular Dynamics Simulations of Whole Viruses. *J. Phys. Chem. Lett*, **9**, 5805–5809. (doi:10.1021/acs.jpclett.8b02298)
6. Tarasova E, Farafonov V, Khayat R, Okimoto N, Komatsu T, Taiji M, Nerukh D. 2017 All-atom molecular dynamics simulations of entire virus capsid reveal the role of ion distribution in capsid's stability. *J. Phys. Chem. Lett.* **8**, 779–784. (doi:10.1021/acs.jpclett.6b02759)
7. Tarasova E, Korotkin I, Farafonov V, Karabasov S, Nerukh D. 2017 Complete virus capsid at all-atom resolution: Simulations using molecular dynamics and hybrid molecular dynamics/hydrodynamics methods reveal semipermeable membrane function. *J. Mol. Liq.*, **245 (Supplement C)**, 109–114. (doi:10.1021/acs.jpclett.6b02759)
8. Tarasova E, Farafonov V, Taiji M, Nerukh D. 2018 Details of charge distribution in stable viral capsid. *J. Mol. Liq.* **265**, 585–591. (doi:10.1016/j.molliq.2018.06.019)
9. Kovacs EW, Hooker JM, Romanini DW, Holder PG, Berry KE, Francis MB. 2007 Dual-Surface-Modified Bacteriophage MS2 as an Ideal Scaffold for a Viral Capsid-Based Drug Delivery System. *Bioconjugate Chem.*, **18**, 1140–1147. (doi:10.1021/bc070006e)
10. Licis N, van Duin J, Balklava Z, Berzins V. 1998 Long-range translational coupling in single-stranded RNA bacteriophages: an evolutionary analysis. *Nucl. Acids Res.*, **26**, 3242–3246.
11. Licis N, Balklava Z, van Duin J. 2000 Forced retroevolution of an RNA bacteriophage. *Virology*, **271**, 298–306. (doi:10.1006/viro.2000.0327)
12. Koning RI, Gomez-Blanco J, Akopjana I, Vargas J, Kazaks A, Tars K, Carazo JM, Koster AJ. 2016 Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. *Nat. Commun.*, **7**, 12524. (doi:10.1038/ncomms12524)
13. Dai X, Li Z, Lai M, Shu S, Du Y, Zhou ZH, Sun R. **2017** In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature*, **541**, 112–116.
14. Golmohammadi R, Valegard K, Fridborg K, Liljas L. 1993 The refined structure of bacteriophage MS2 at 2.8A resolution. *J. Mol. Biol.* **234**, 620–639. (doi:10.1006/jmbi.1993.1616)
15. James M, Murtola T, Schulz R, Smith JC, Hess B, Lindahl E. 2015 GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *ScienceDirect*, **2**, 19–25. (doi:10.1016/j.softx.2015.06.001)
16. Kabsch W, Sander C. 1983 Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637. (doi:10.1002/bip.360221211)
17. Perkett MR, Mirijanian DT, Hagan MF. 2016 The Allosteric Switching Mechanism in Bacteriophage MS2. *J. Chem. Phys.* **35101**, 1–15. (doi:10.1063/1.4955187)
18. Armanious A, Aeppli M, Jacak R, Refardt D, se Sigstam T, Kohn T, Sander M. 2015 Viruses at Solid−Water Interfaces: A Systematic Assessment of Interactions Driving *Adsorption. Environ. Sci. Technol.* **50**, 732−743. (doi:10.1021/acs.est.5b04644

19. Shao F. 2014 Toxicity of silver nanoparticles on virus. *J. Mater. Environ. Sci.* **5**, 587-590.
20. Michen B, Graule T. 2010 Isoelectric points of viruses. *J. Appl. Microbiol.* **109**, 388-97. (doi:10.1111/j.1365-2672.2010.04663.x)