The structure and physical properties of a packaged bacteriophage particle

https://doi.org/10.1038/s41586-024-07150-4

Received: 1 September 2023

Accepted: 1 February 2024

Kush Coshic[1], Christopher Maffeo[2,3], David Winogradoff[3] & Aleksei Aksimentiev[1,2,3] ✉

A string of nucleotides confined within a protein capsid contains all the instructions necessary to make a functional virus particle, a virion. Although the structure of the protein capsid is known for many virus species[1,2], the three-dimensional organization of viral genomes has mostly eluded experimental probes[3,4]. Here we report all-atom structural models of an HK97 virion[5], including its entire 39,732 base pair genome, obtained through multiresolution simulations. Mimicking the action of a packaging motor[6], the genome was gradually loaded into the capsid. The structure of the packaged capsid was then refined through simulations of increasing resolution, which produced a 26 million atom model of the complete virion, including water and ions confined within the capsid. DNA packaging occurs through a loop extrusion mechanism[7] that produces globally different configurations of the packaged genome and gives each viral particle individual traits. Multiple microsecond-long all-atom simulations characterized the effect of the packaged genome on capsid structure, internal pressure, electrostatics and diffusion of water, ions and DNA, and revealed the structural imprints of the capsid onto the genome. Our approach can be generalized to obtain complete all-atom structural models of other virus species, thereby potentially revealing new drug targets at the genome–capsid interface.

The essence of each virus, its genome, is protected from the outside world by a protein enclosure, a capsid. Once conditions are met for viral invasion, the genome is released from the capsid into a target cell. Advances in X-ray crystallography and cryogenic electron microscopy (cryo-EM) have provided a wealth of structural information about viral proteins, including fully atomistic structures of assembled capsids for many virus species[1,2]. Much less is known about the global organization of viral genomes inside fully packaged viral capsids[1,2,8].

Genome packaging has been extensively studied in several viral systems[6,9] and has led to characterization of the structural[8,10,11] and mechanical[12,13] aspects of the packaging process. Viral capsids can self-assemble around their genomes, such as in most RNA viruses[3,14]. Alternatively, the genome is packaged into a pre-formed, immature capsid, such as in bacteriophages, including HK97 (ref. 5), which is the focus of this work. Encapsulation of genomes through self-assembly usually requires specific genome–capsid contacts, which has enabled some experiments to resolve as much as 80% of the backbone structure of the genome[3,4]. Precise structural organization of the genome is less understood in pressurized viruses. Small-angle X-ray scattering (SAXS) has enabled the determination of the average inter-DNA distance in packaged viruses[15,16]. Meanwhile, cryo-EM microscopy has shown that DNA forms coaxial shells of increasing disorder towards the interior of the capsid[9,17,18]. Experimentally resolving the genome structure of an individual virion with single-nucleotide resolution is currently not possible.

Modelling the structural organization of nucleic acids inside viral particles is challenging because of the complexity of the ion atmosphere[19,20], scarce experimental data on the energetics of such systems[21,22], the size (50–100 nm diameter) of typical virions and an astronomical number of possible nucleic acid conformations that satisfy lax experimental constraints. Coarse-grained (CG) models[23], which trade precision for performance, have enabled computer simulations of entire virions, including SARS-CoV-2 (ref. 24) and HIV[25]. Packaging of DNA in bacteriophage systems has been investigated using several CG models[26–30], which, under specific assumptions regarding DNA–capsid[28,29] or DNA–DNA[30] interactions, could produce coaxial spooling of the genomes. Despite offering an accurate description of DNA mechanics[31,32], DNA–DNA and DNA–protein interactions[33], all-atom molecular dynamics (MD) simulations of the genome packaging process have not yet been possible. Consequently, previous all-atom MD studies of viral systems[34] have analysed the behaviour of protein capsids[35,36] or outer envelopes[37,38], leaving the structure and dynamics of the nucleic acids confined within the capsids largely unexplored, with the exception of a proof-of-principle study[39].

Here we report complete all-atom structural models of a bacteriophage HK97 virion obtained through a multiresolution simulation approach. Using a coarser-resolution model, we package the entire 39,732 base pair (bp) genome of HK97 into an atomically accurate representation of its 65 nm diameter capsid[9,15,17]. By gradually increasing the resolution of the computational model, we arrive at a fully atomistic, explicit solvent representation of the HK97 virion, which we then simulate using the all-atom MD method. Analysis of our multiresolution simulations provides a wealth of information not easily accessible by other means.

---

[1]Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [2]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [3]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL, USA. ✉e-mail: aksiment@illinois.edu

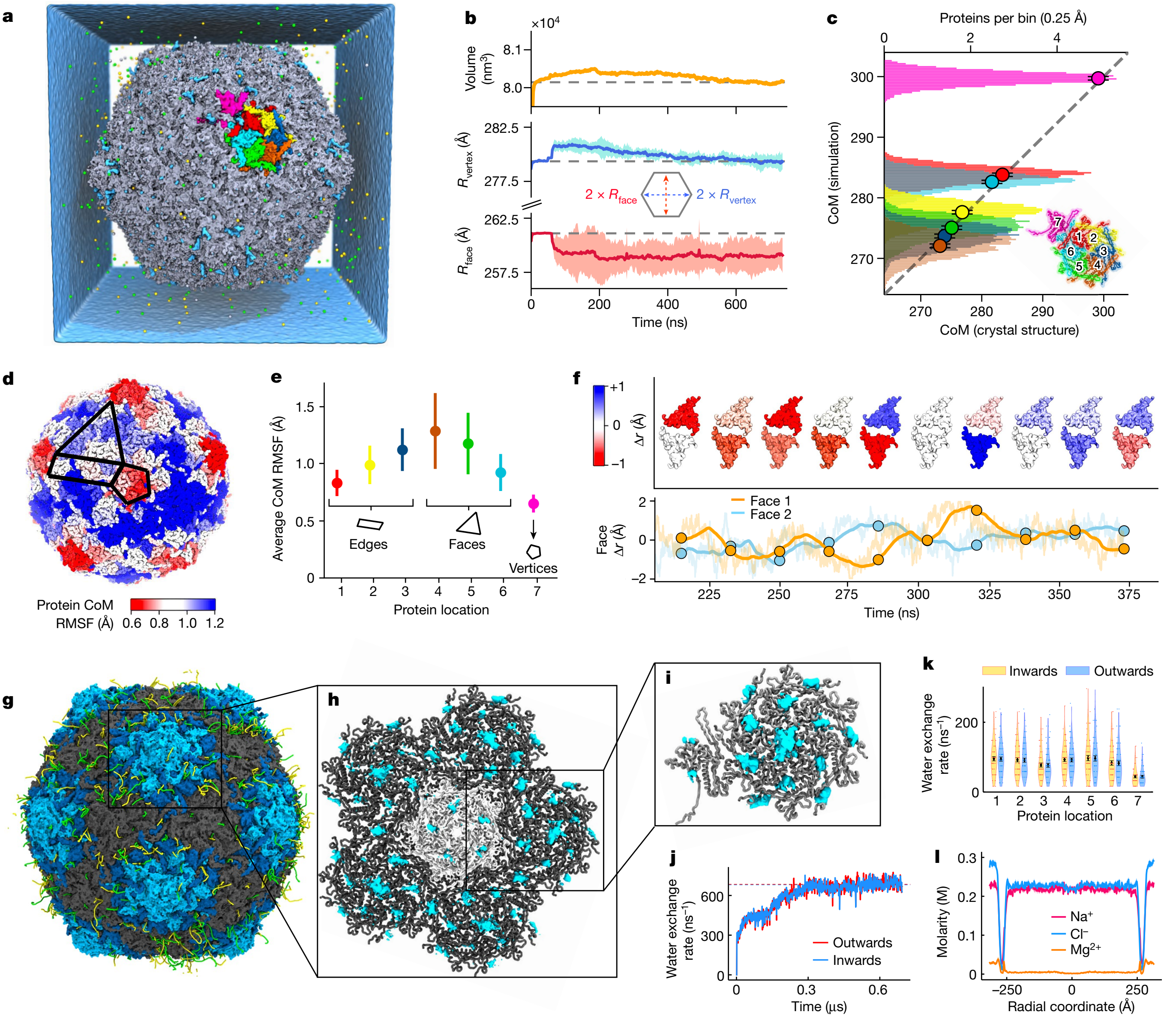


Fig. 1 | In situ structure of a protein capsid in the absence of DNA. **a**, Solvated, 27.5 million atom model of an empty HK97 virion. One of the repeating subunits of the icosahedral capsid is highlighted in colours that represent the seven protein locations within the subunit. **b**, Interior volume (top), vertex-to-vertex (middle) and face-to-face (bottom) dimensions of the capsid versus simulation time. The dashed lines indicate the corresponding crystal structure values. Error bars show the s.d. over $n = 6$ and $n = 10$ pairs of vertices and faces, respectively. **c**, Trajectory-average distance from the centre of the capsid to the CoM of each capsid protein versus the corresponding crystal structure value. For each protein location, the symbol shows the value averaged over the 60 copies of the protein, whereas the histogram (top axis) shows the distribution of the value among the copies. **d**, Root mean-squared fluctuation (RMSF) of each protein CoM coordinate over a representative 50 ns fragment of equilibration. **e**, Average CoM RMSF by protein location. Error bars represent the s.d. over 60 copies per protein. Edge, face and vertex parts of the capsid are defined in **d** and **e**. **f**, CoM radial displacement of two neighbouring faces with respect to the average value shown using a colour bar (top) and as a 10 ns running average (bottom). **g**, Representative 1 ns trajectories of water molecules crossing into (green) and out of (yellow) the capsid. **h**,**i**, Location of water passages (blue) visualized as an isosurface (5.26 molecules per nm^3) of water density. For **i**, the density was averaged over the 60 icosahedron subunits. **j**, Trans-capsid water exchange rate versus simulation time. **k**, Equilibrium water exchange rate for each protein location. Error bars (black) represent the s.e.m. over 60 copies per protein. **l**, Equilibrium ion concentration versus distance from the capsid centre.

In situ properties of an empty capsid

As a baseline reference for our computational models of fully packaged virions, we determined the in situ structure and physical properties of an empty, fully mature HK97 capsid (Fig. 1a). The initial all-atom model of the capsid was constructed starting from its crystal structure (Protein Data Bank (PDB) identifier: 2FT1)[40]. The titratable residues at the inner surface of the capsid were protonated[41], which made the capsid electrically neutral. After solvation with 0.2 mol kg^{-1} NaCl and 0.005 mol kg^{-1} $MgCl_2$ electrolyte, minimization and restrained equilibration, the capsid was simulated for 0.7 μs without restraints (Methods).

The overall structure of the capsid was observed to attain an equilibrium configuration after approximately 0.5 μs. After initial increase, the volume enclosed by the capsid (Fig. 1b, top), and the average location of the vertices of the capsid (Fig. 1b, middle), returned to their crystal structure values, whereas the faces of the capsid moved inwards by 2 Å (Fig. 1b, bottom). The radial distance to individual gp5 proteins deviated from its crystal structure value, but remained within around 1 Å of the latter (Fig. 1c). In contrast to a previous simulation of the HIV capsid[35], the root mean square deviation (RMSD) of resolved residues in the HK97 capsid converged below 3.5 Å within the initial 0.3 μs (Extended Data Fig. 1a,b). Among the seven protein locations constituting an asymmetric subunit (Fig. 1c), proteins forming the vertex (location 7) showed the greatest deviation from the crystal structure and the least amount of fluctuation (Extended Data Fig. 1c–f).

During equilibration, the capsid was seen to undergo rippling fluctuations; that is, alternating outward and inward radial movements of the capsid subunits (Supplementary Video 1). The amplitude of such fluctuations depended on the location within the repeating structural unit of the capsid (Fig. 1d), with the centre of mass (CoM) of proteins at the faces and edges fluctuating significantly more than those at the vertices (Fig. 1e). Figure 1f illustrates the rippling motion by plotting the radial displacements of a pair of neighbouring faces, such that the inward motion of one face is anticorrelated with an outward motion of the other. On a larger scale, the rippling motion was not coherent, and its period varied from one pair of faces to the other. Simultaneously, the capsid underwent global fluctuations that involved expansion and contraction of the capsid as a whole, and resembled the act of 'breathing'. Principal component analysis revealed the top three modes of the collective motion that describe asymmetric deformation of the capsid featuring opposite direction motion of its nearby regions (Extended Data Fig. 1g–h and Supplementary Video 2). These modes were similar to the behaviour previously reported from simulations of other empty viral capsids[36,42].

Over the course of the equilibration simulation, water molecules were observed to pass into and out of the capsid (Fig. 1g–i), with the rate of water exchange increasing to over 600 molecules per ns in equilibrium (Fig. 1j). This equilibrium exchange rate is higher than that reported for a poliovirus[43] (around 10 ns^{-1}), but smaller than the exchange rate through a hepatitis B capsid[36] (about 10^3 ns^{-1}) or an HIV capsid[35] (about 10^4 ns^{-1}). The water exchange was not uniform along the capsid surface (Fig. 1k), with the majority of the water transport occurring at the faces and edges of the capsid and not at the vertices, despite each of the latter having a pore (Fig. 1h). Sodium and chloride ions also permeated through the capsid, reaching an equilibrium exchange state after 0.4 μs (Extended Data Fig. 2a). At equilibrium, chloride ions crossed the capsid 15 times more frequently than sodium ions. At the timescale of our simulation, no crossing events were observed for magnesium ions, which were localized near the junction of the asymmetric subunits. It remains unclear whether the treatment of magnesium ions as permanent hexahydrate complexes—bulkier than bare magnesium but less prone to forming contacts that are irreversible on the MD timescale—is responsible for this observation.

Genome packaging and virion individuality

Using our multiresolution model of DNA, parametrized to reproduce experimentally measured DNA elasticity and osmotic pressure in condensed DNA arrays, we simulated packaging of the 39,732 bp genome into an implicit, grid-based representation of the fully mature HK97 capsid, bypassing the maturation process and the associated capsid expansion (see Methods for details). Starting with a well-equilibrated DNA conformation (Fig. 2a), DNA packaging was produced by applying a local force to the DNA beads located within the portal of the capsid (Supplementary Video 3). Sixteen packaging simulations were performed at 4 bp per 2 bead resolution (1 bead representing the helix and 1 bead representing the orientation of the major groove; implicit solvent) and a packaging force of 55 pN (referred to as 'slow'; Extended Data Fig. 3 and Supplementary Video 4). In eight of the simulations, a local torque was applied in the portal to rotate the DNA right-handed by 14° per 1 turn of the packaged DNA[44]. After packaging, each system was equilibrated for approximately 200 μs, which produced 16 distinct configurations of a fully packaged virion. Four additional packaging simulations were performed at 4 bp per 2 bead resolution and a higher packaging force of 280 pN (referred to as 'fast'), and for 2 of these, a local torque was applied to rotate the DNA by around 360° for each turn of the packaged DNA. Note that the genome packaging speed in the slow packaging simulations approaches the speed of unassisted genome ejection from the capsid (Fig. 2a and Supplementary Video 5), with the ratio of genome packaging-to-genome ejection speeds congruent with the experiment[45].

During the simulations, the fraction of packaged genome steadily increased (Fig. 2a) and exhibited small fluctuations, which were caused by the differences in the DNA configuration outside the capsid. The pressure inside the capsid monotonically increased, reaching approximately 10 bar for a fully packaged configuration (Fig. 2b). The bend energy per packaged base pair decreased slightly as packaging progressed (Fig. 2b). This result is apparently at odds with the spooling model of DNA packaging[26,28], whereby early packaged DNA forms gently curved concentric rings at the edges of the capsid. In a previous simulation of DNA spool relaxation[46], large-scale rearrangement of the inner DNA spools were observed to reduce the bending stress. In our simulations of DNA packaging, a markedly different mechanism for reducing global stress was seen to emerge. Close inspection of the packaging simulations revealed a repeating loop extrusion process that begins with a highly curved bulge forming near the portal (Fig. 2c). The incoming DNA then pushes the bulge away from the portal to form a hairpin-like switchback loop that propagates along an established local helical axis until sufficient stress builds that a new bulge forms at the portal and the process is repeated (Supplementary Video 6). The average length of the loop gradually increased from 50 to 150 bp as the genome filled the capsid (Fig. 2d). By the end of the simulations, about three quarters of the genome formed such loops (Fig. 2d).

The switchback-loop organization of the DNA genome was partially caused by the significant local liquid crystalline (nematic) order, which developed early during the packaging process (Fig. 2e). In a fully packaged particle, the nematic order was the highest in the outermost parts of the genome (Extended Data Fig. 3b), which could be a consequence of the capsid geometry preventing the helices from adopting sharply radial orientations at the boundary. Alternatively, it is possible that the non-uniform nematic order is indicative of a slow relaxation process[47] that prevents full equilibration of the nematic order throughout the system. By interpolating both the nematic axis and the nematic order to a regular grid and using theoretical values for the elastic constants[48], we numerically computed the local Frank–Oseen energy that quantifies the cost of introducing distortions into a nematic field (Fig. 2e). Dominated by the splay term, this energy monotonically increased during packaging. As most of the packaged genomes visually appeared to exhibit significant order about the axis of the packaging portal, we expected to find a high value for the toroidal order parameter[27], a metric that is 0.5 for a random coil and 1.0 for a perfectly spooled polymer. The toroidal order reached around 0.6 at the end of the packaging simulation (Fig. 2f). This low value can be attributed to microscopic deviation of the DNA path out of the plane defined by the toroidal axis. The initially high value of the toroidal order can be attributed to our choice to compute the toroidal axis from the DNA path, adopting the axis that maximizes the toroidal order rather than the portal axis. A detailed analysis of the fully equilibrated slow packaged configurations showed little dependence of the global genome metrics on the presence of a twist during the packaging process (Extended Data Fig. 3b).

Although the packaging simulations started from similar initial configurations of the genome, and although many statistical measures exhibited low variance across packaged capsids (Fig. 2b–f and Extended Data Fig. 3b), the heterogeneity of the packaged structures was visually striking (Supplementary Video 4 and Extended Data Fig. 3a). For example, a sharp interface was observed in the outmost packaged DNA between early-packaged and late-packaged DNA, and the contour of

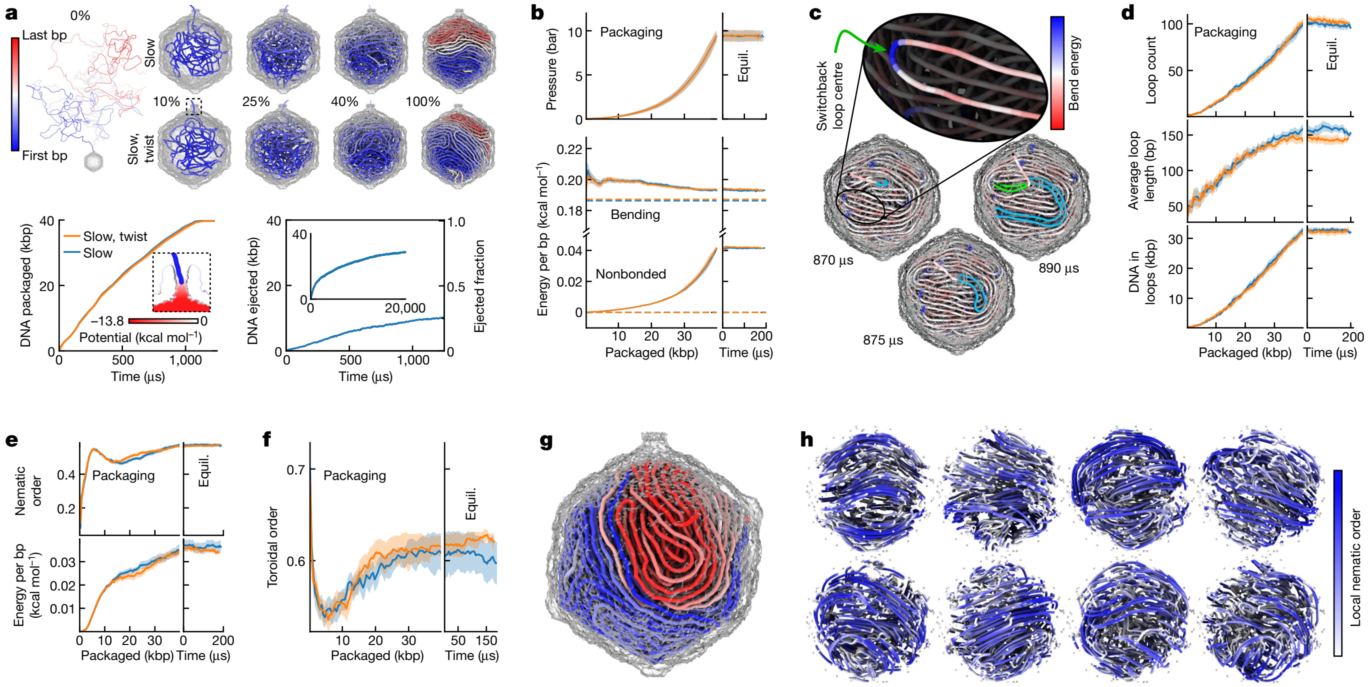


Fig. 2 | Simulation of genome packaging. a, Top, packaging of the HK97 genome at 4 bp per 2 bead resolution. Bottom, the plots illustrate simulations of the packaging process (left) driven by the portal potential (inset) and of spontaneous ejection (right). Solid lines here and throughout the figure depict ensemble averages; shaded regions depict the s.e.m. among eight independent packaging simulations (replicas) for each curve. **b**, Internal pressure and energy during packaging and equilibration (Equil.) simulations. **c**, Switchback-loop formation during packaging. Inset depicts a switchback loop with a highly curved centre and two arms that remain within 6 nm. The three genome configurations depict the formation of two switchback loops (in cyan and green). The first switchback loop (cyan) is extruded by the packaging motor until growth of the loop stalls and the nascently packaged DNA buckles, forming a second loop (green). The DNA is coloured by the instantaneous bend energy from 0 kcal mol^{-1} (red) to 1.5 kcal mol^{-1} (blue). **d**, Quantification of switchback loops during packaging, including the absolute number of loops detected, the average loop length (note that loops may overlap) and the amount of the packaged DNA in at least one switchback loop. **e**, Global order in the packaged genome. The local nematic order (6 nm neighbourhood), and the Frank–Oseen nematic energy calculated from the nematic director field. **f**, Toroidal order of the packaged genome. **g**, Example of a packaged genome featuring a baseball seam interface between early-packaged (blue) and late-packaged (red) DNA. **h**, Streamlines of the nematic director field of select capsids reveal a distinct pattern for each genome. Each streamline is coloured by the local nematic order with white corresponding to zero order, and blue corresponding to the maximum order observed in the system.

this interface varied, being roughly planar in some capsids (Fig. 2a) and baseball seam shaped in others (Fig. 2g). Furthermore, the planar interfaces formed at various angles relative to the packaging axis. Independent characterization of the four packaged genome quadrants produced highly similar results. On average, the most notable difference was a few per cent increase of the local nematic order for the last quadrant to be packaged. To visualize the differences among the capsids, the nematic director field was treated like a velocity field through which streamlines could be drawn (Fig. 2h). The streamlines showed that the genome of each capsid possesses a distinct liquid crystal order[49], despite having similar global properties, including pressure, energy and average nematic order parameter. Thus, the packaging process gives mature virions individual traits encoded by the topology of the packaged genome.

Properties of a packaged capsid

We built our initial all-atom models of fully packaged virions by first converting the equilibrated CG models of packaged DNA to higher resolution CG models (2 bp per 2 beads and then to 1 bp per 2 beads), equilibrating each model for 400 µs. Six systems that differed by the configuration of the packaged DNA were then converted to an all-atom representation that had the sequence of HK97 genome and merged with a pre-equilibrated volume of electrolyte with the ion composition of a tight DNA bundle[50]. Solvent molecules that clashed sterically with the genome were removed. The resulting systems were equilibrated in the presence of a grid-based potential that reproduced the confinement of the protein capsid while allowing free exchange of water and ions. After the ion atmosphere reached equilibrium, the grid potential was replaced by the all-atom structure of the capsid. The resulting systems underwent multiple steps of restrained equilibrations and were eventually simulated in the absence of any restraints (Methods and Supplementary Tables 1–9). Four models containing DNA packaged at a high packaging force with and without twist (2 models each) were each simulated for 0.31 to 0.87 µs, depending on the system. Two models containing DNA packaged at a low packaging force with and without twist were each simulated for about 1 µs. Supplementary Video 7 illustrates one such equilibration trajectory, and Fig. 3a shows its final equilibrated state.

Free equilibration simulations of the fully packaged virions revealed a balloon-like expansion of the virion volume manifested by an outward motion of the faces of the capsids and a near constant separation of the vertices of the capsids (Fig. 3b and Supplementary Video 1). In all simulations of the packaged capsids, the 6 proteins forming the faces of the capsid moved by 3–5 Å outwards, whereas the proteins at the vertices moved less than 1 Å (Fig. 3c and Extended Data Fig. 4a). The rippling motion of the packaged capsids was decreased (Supplementary Video 1) and the average CoM fluctuation of the proteins was reduced by half (Fig. 3d). The 5% expansion of the capsid volume matched the 4% increase of the volume occupied by the genome. Thus, the self-repulsion of the DNA genome stretches the capsid while simultaneously reducing the amplitude of its local fluctuations.

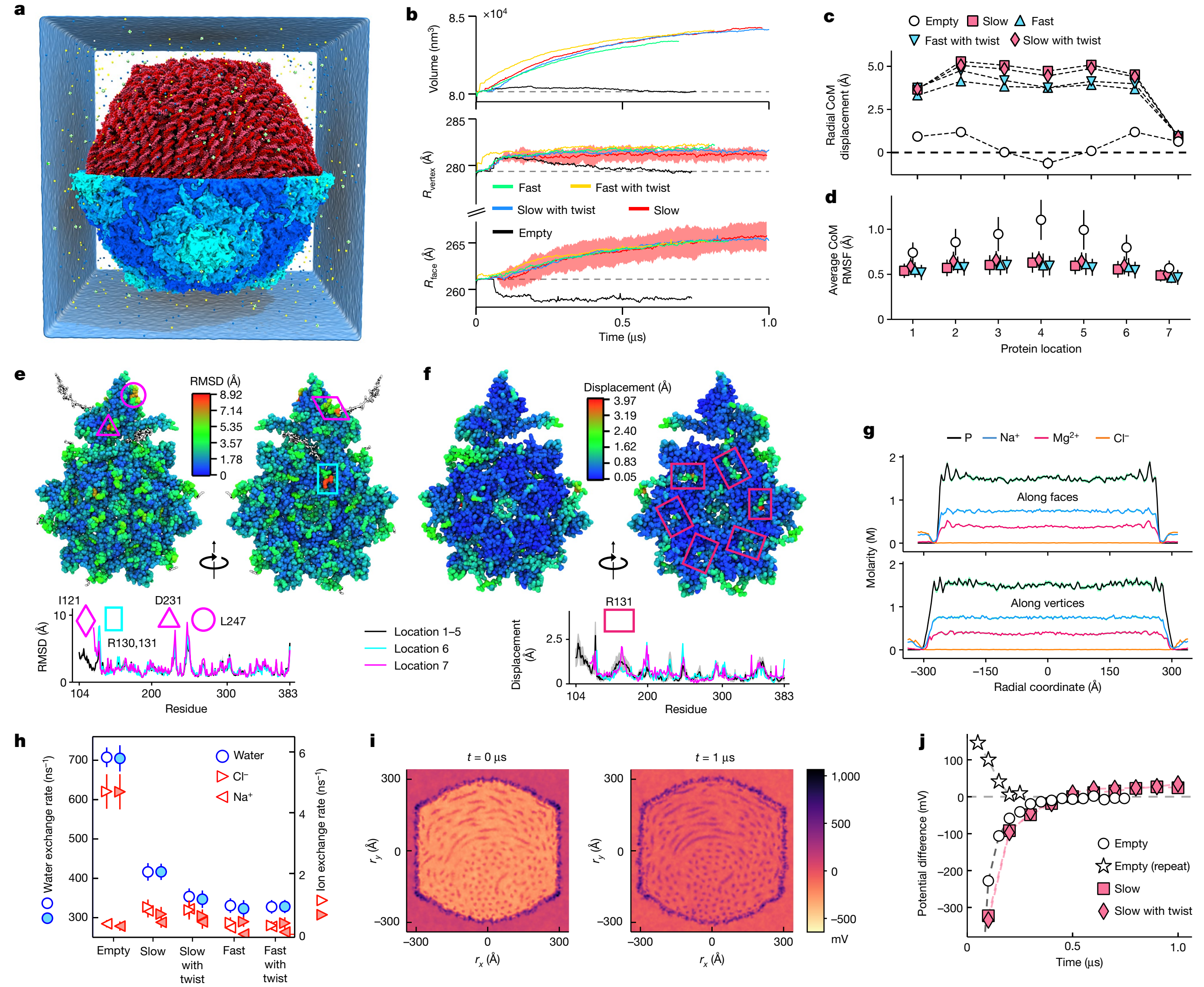


Fig. 3 | Properties of a packaged capsid. a, Solvated, 26 million atom model of a packaged virion after 1 µs of MD simulation. To reveal the encapsulated DNA (red), one half of the protein capsid is not shown. **b**, Interior volume (top), vertex-to-vertex (middle) and face-to-face (bottom) dimensions of the capsid simulated with and without DNA. Error bars show the s.d. over $n = 6$ and $n = 10$ pairs of vertices and faces, respectively. **c**, Average CoM displacement of the capsid proteins from their crystallographic coordinates. **d**, Average CoM RMSF of the protein capsid by the protein location within the asymmetric subunit. Error bars represent the s.d. over 60 copies of the protein. **e**, Average per-residue RMSD of a fully packaged capsid (slow trajectory) depicted by colouring the outward-facing (top left) and inward-facing (top right) residues of the icosahedron unit and as a function of the residue number (bottom). Residues undergoing large deviations are marked with symbols coloured according to the protein location. Residues not resolved in the crystal structure are shown in white. **f**, Average per-residue displacement of the packaged capsid with respect to the empty capsid. **g**, Equilibrium ion concentration versus distance from the capsid centre. **h**, Equilibrium rate of water exchange across the capsid, inwards (open) and outwards (solid), for fully packaged and empty states of the capsid. Error bars show the s.d. over $n = 25$ measurements. **i**, Electrostatic potential map of the packaged virion (slow) averaged over the first (left) and the last (right) 100 ns fragments of the equilibration simulation. **j**, Average electrostatic potential of the volume enclosed by the capsid relative to that outside the capsid as a function of simulation time for empty and packaged capsid simulations.

The RMSD of individual residues from their initial coordinates provided a detailed map of the microscopic deformations (Fig. 3e). Higher than average displacements were observed for residues at the boundaries of neighbouring capsid proteins and at the vertices, which correlated with the locations where solvent exchange occurs. Conversely, lower than average displacements were seen for residues buried under the protein surface. Per residue comparison of the empty and packaged structures (Fig. 3f) revealed modest differences, which suggested that large (>7 Å RMSD) deviations from the crystal structure coordinates (Fig. 3e) are not caused by the pressure from the packaged DNA. By contrast, larger deviations of certain residues, such as of Arg130, Arg131 and the adjacent residues, can be attributed to their specific interactions with the packaged genome. Similar structural deformations of the capsid proteins were observed in other simulations of the packaged particles (Extended Data Fig. 4).

The presence of DNA makes the ion composition inside the capsid different from the outside. The negative charge of the DNA phosphate groups was compensated approximately in equal amounts by the increased concentrations of Na^+ and Mg^{2+} cations, whereas Cl^- ions are depleted[50] (Fig. 3g). The charge of the DNA also polarized the capsid, increasing the charge of its inward facing surface from 81 ± 10 to 305 ± 10 proton charges. As with the empty capsid, solvent exchange occurred primarily through the small gaps at the protein interfaces, although the equilibrium rate of water exchange was reduced by a factor of two (Fig. 3h). This reduction was caused by a 30% (by volume) contraction of water-filled cavities (Extended Data Fig. 5a) and reduced water diffusivity (Fig. 4j), which also accounted for the slight dependence of the rate on the packaged model. The rate of water exchange across the capsid was more than 100 times higher than the rate of net water flux into the capsid during equilibration, which increased solvation of the genome by 0.6%. The DNA reduced the rate of Cl^- exchange by more than fivefold, whereas the rate of Na^+ exchange increased by a factor of two (Extended Data Fig. 2a), which qualitatively matched the changes in the respective ion concentrations.

The average maps of the electrostatic potential (Fig. 3i and Extended Data Fig. 2d–f) revealed substantial changes in the interior of the packaged capsid. Initially, the average potential of the interior of the capsid was approximately −350 mV relative to that of the outside volume (Fig. 3j and Supplementary Video 8). The internal potential reached 0 mV after 500 ns of equilibration but continued to increase, saturating at 25 ± 2 mV. A similar change in the electrostatic potential was observed during a microsecond-long equilibration of another model of the packaged capsid (Fig. 3j). By contrast, the interior electrostatic potential of an empty capsid equilibrated to zero, which we confirmed by repeating empty capsid equilibration using different initial ion concentrations (Fig. 3j). The gradual increase in the interior electrostatic potential is produced by the influx of ions neutralizing the electrical charge of the DNA (Extended Data Fig. 2d), whereas the overall positive value of the equilibrated potential is caused by the inner core of the DNA (Extended Data Fig. 2h). Overall, a packaged capsid particle has an electrically neutral core surrounded by a polarized protein shell (Extended Data Fig. 2e–h).

Properties of packaged DNA

Multiparticle cryo-EM reconstructions of DNA bacteriophages, including HK97 (ref. 17), showed distinct coaxial layers of DNA density (Fig. 4a). To determine whether individual genome configurations obtained from our DNA packaging simulations are consistent with such averaged cryo-EM density, we averaged the DNA density sampled by individual MD runs over the icosahedral symmetries of the capsid and the MD trajectory. The resulting average density (Fig. 4b) showed multiple distinct coaxial layers that diminished in intensity towards the centre of the capsid. The plots of the averaged experimental and simulated DNA densities across the capsid (Fig. 4c) revealed five (along faces) and seven (along vertices) peaks. The average simulated DNA–DNA spacing was 25.3 ± 0.2 Å along faces and 26.1 ± 1.4 Å along vertices, whereas the average inter-axial distance computed directly[51] from the 3D genome configuration was 25.3 ± 0.5 Å. Similar layering of averaged densities was observed for other equilibrated all-atom models (Extended Data Fig. 7a).

Despite the emergence of coaxial layers in the average DNA density, the instantaneous genome configurations produced by our packaging simulations (Fig. 2 and Extended Data Fig. 3) did not conform to a perfect coaxial inverse-spool, as seen in previous modelling studies of the T7 system[28] or the ε15 system[29]. To demonstrate how coaxial spooling emerges from multiparticle averaging, we applied our averaging procedure to our CG equilibration trajectories of the packaged particles. As we increased the number of symmetries used for averaging (Fig. 4d), an increasing number of layers became visible. By averaging over the 16 independent packaged configurations, 7 coaxial layers were clearly identified, similar to that seen in an experimental reconstruction[17]. Diffuse spool-like arrangement of DNA near the capsid was also visible in some single-particle transmission electron microscopy (TEM)-like images of the simulated DNA configurations (Extended Data Fig. 6), which stems from projection averaging inherent to the TEM method. Thus, the presence of layers in the averaged structures may not be specifically interpreted as an indication of the inverse-spool-like arrangement of the packaged DNA, which echoes conjectures from previous studies[27,49,51,52].

The global organization of the genome was further assessed by calculating SAXS profiles separately for the capsid and the DNA parts of the packaged particle. Similar to the experimental SAXS profile of an empty HK97 capsid[15], the calculated profiles of the packaged capsid featured multiple peaks at longer wavelengths reflecting the icosahedral symmetry (Fig. 4e). Over the 1 μs simulation, the SAXS profiles of the genome converged to a shape characterized by the DNA diffraction peak at 0.248 Å^{-1} (Fig. 4e, right), a result also seen in all equilibrated packaged structures, in agreement with the 0.238–0.264 Å^{-1} range reported in previous studies[53,54].

The overall structure of the genome appeared less ordered towards the capsid centre, which we characterized by computing the angle between the local helical axis of the DNA and its radial vector from the centre of the capsid. The resulting distributions (Fig. 4f and Extended Data Fig. 7e) showed that DNA helices are almost always perpendicular to their radial vectors near the capsid but lose that orientation halfway towards the capsid centre. At the base-pair level, 3 out of 12 bp characteristics showed a progressive change with distance from the capsid centre (Fig. 4g and Extended Data Fig. 8). In proximity to the capsid, DNA base pairs had stagger similar to that of regular B-DNA, whereas the opening and stretch parameters deviated appreciably from the B-DNA values. Moving towards the interior, the absolute value of stagger increased as the stretch parameter returned to its B-DNA value. The fraction of base pairs that deviated appreciably from their Watson–Crick configurations (Fig. 4h and Extended Data Fig. 7b,c) reached a maximum at the outermost layer of the DNA, a result consistent with a picture in which the stress on the genome increases with the distance from the capsid centre. These broken base pairs could be broadly categorized as mis-stacked, frayed and over-bent (Extended Data Fig. 7d). Mis-stacking accounted for around 50% of the population, and the proportion of over-bent configurations was higher in the capsid interior.

To characterize the local mobility of the genome inside the packaged capsid, we computed local diffusion constants of each 10 bp fragment of the packaged DNA (Fig. 4i, right). The distribution showed two populations, which, by plotting the radial dependence of the diffusion constant (Fig. 4i, left), we associated with DNA fragments being away from and in proximity to the protein capsid.

Local displacements of the DNA were analysed both in the reference frame of the capsid and the local helical frame (Extended Data Fig. 5b). Radial and longitudinal mobility of the DNA were similar relative to the capsid; however, when viewed in the local helical frame, the DNA was more mobile in the direction transverse to its local helical axis in comparison to that along the axis, opposite to the situation in straight DNA bundles[50]. The computed diffusion coefficients were similar to those ($3\ \mu m^2\ s^{-1}$) experimentally measured using a fluorescently labelled DNA packaged into a T4 capsid[55]. The presence of packaged DNA reduced the local diffusivity of water by 2.75 times (Fig. 4j). Notably, a similar reduction was observed for Cl^- ions (Extended Data Fig. 5c), whereas diffusion of Na^+ and Mg^{2+} was reduced by factors of four and eight, respectively, which qualitatively resembles ion localization effects predicted to occur in dense DNA systems[20].

Protein–DNA interactions

In a packaged particle, self-repulsion of the packaged DNA compels it to expand radially outwards. This expansion is confined by the protein capsid, of which the icosahedral symmetry may imprint distinct structural features onto the outer layer of the packaged genome. In search of such structural features, we visualized the outermost layer of the packaged DNA, categorizing its parts according to their proximity to the faces, edges and vertices of the capsid (Fig. 5a).

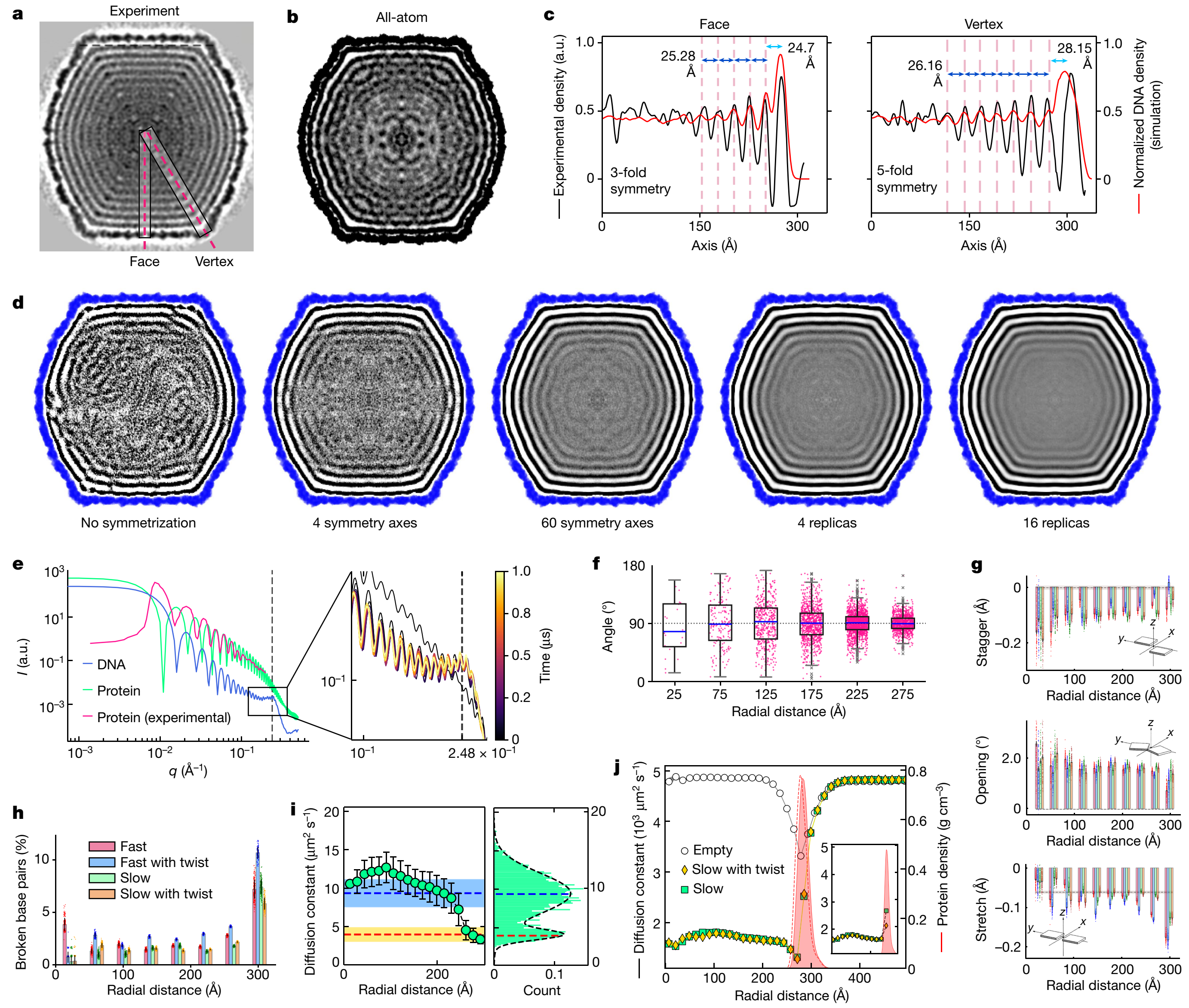


Fig. 4 | Properties of packaged DNA. a,**b**, Average density of packaged DNA determined by cryo-EM[17] (**a**) and cryo-EM-like analysis of the all-atom MD trajectory (**b**). **c**, Simulated (right axis) and experimental (left axis; in arbitrary units (a.u.)) profiles of DNA density. **d**, Cryo-EM-like density analysis of DNA configurations sampled by CG simulations of packaged particles shown, from left to right, with increasing degree of averaging. Replicas refer to independent packaged configurations. **e**, Left, SAXS profiles of the DNA and protein components of the all-atom model and experimental profile of an empty capsid. Dashed line indicates the DNA diffraction peak. Right, SAXS curves for DNA configurations along the slow trajectory. **f**, Box plots of the angle between the local axis of a 10 bp DNA fragment and a radial vector versus distance to the capsid centre. Each box shows the quartiles of the dataset, blue line denotes median and whiskers are drawn at 1.5 times the interquartile range. **g**, Base-pair-level characterization of the packaged genome, normalized by the number of base pairs within each radial bin. Error bars depict the s.e.m. over $n = 104$ consecutive segments (last 5 ns) of the trajectory. Colours, as defined in **h**, indicate packaged models. **h**, Fraction of broken base pairs versus distance from the capsid centre. Error bars depict the s.d. in respective radial bins. **i**, Local diffusion constants of a 10-bp DNA fragment versus its distance from the capsid centre (left) and the distribution of the constant (right). The colour bands indicate the widths of a double Gaussian fit to the histogram. **j**, Local diffusion constant of water (left axis) versus radial distance for empty and packaged capsid. The inset shows the local diffusion constants near and inside the protein capsid. The radially averaged protein density (right axis) is shown for the packaged (filled distribution) and empty (dashed line) trajectory. Image in **a** adapted from ref. 17, Elsevier. Image in **g** adapted from ref. 59, Springer Nature.

Using a line that connected neighbouring vertices to define the orientation of each edge, we found that the outer DNA helices are most likely to cross the capsid edges under a 45° angle (Fig. 5b, top), rarely aligning with the edge or crossing normal to it. The icosahedral confinement of the capsid produces occasional defects[49] in the outer layering of the genome (Fig. 5b, bottom, and Extended Data Fig. 9), which manifest as unoccupied spaces between neighbouring DNA helices. In its entirety, the shape of the packaged DNA deviated significantly from a sphere (Fig. 5c and Supplementary Video 9). The DNA closer to the capsid vertices bulged outwards, protruding directly into the vertices. Defining a coordinate system based on the five proteins that make up each vertex, we observed a slight preference in the orientation of the protruding loop (Fig. 5d). The base pairs of these loops were often distorted or broken, with several instances of non-Watson–Crick base pairs forming sustained contacts with the protein capsid.

The plot of averaged DNA density (Fig. 4c) showed that the distance between the outer DNA layer and the capsid was slightly smaller than the average DNA–DNA distance near the faces of the capsid but

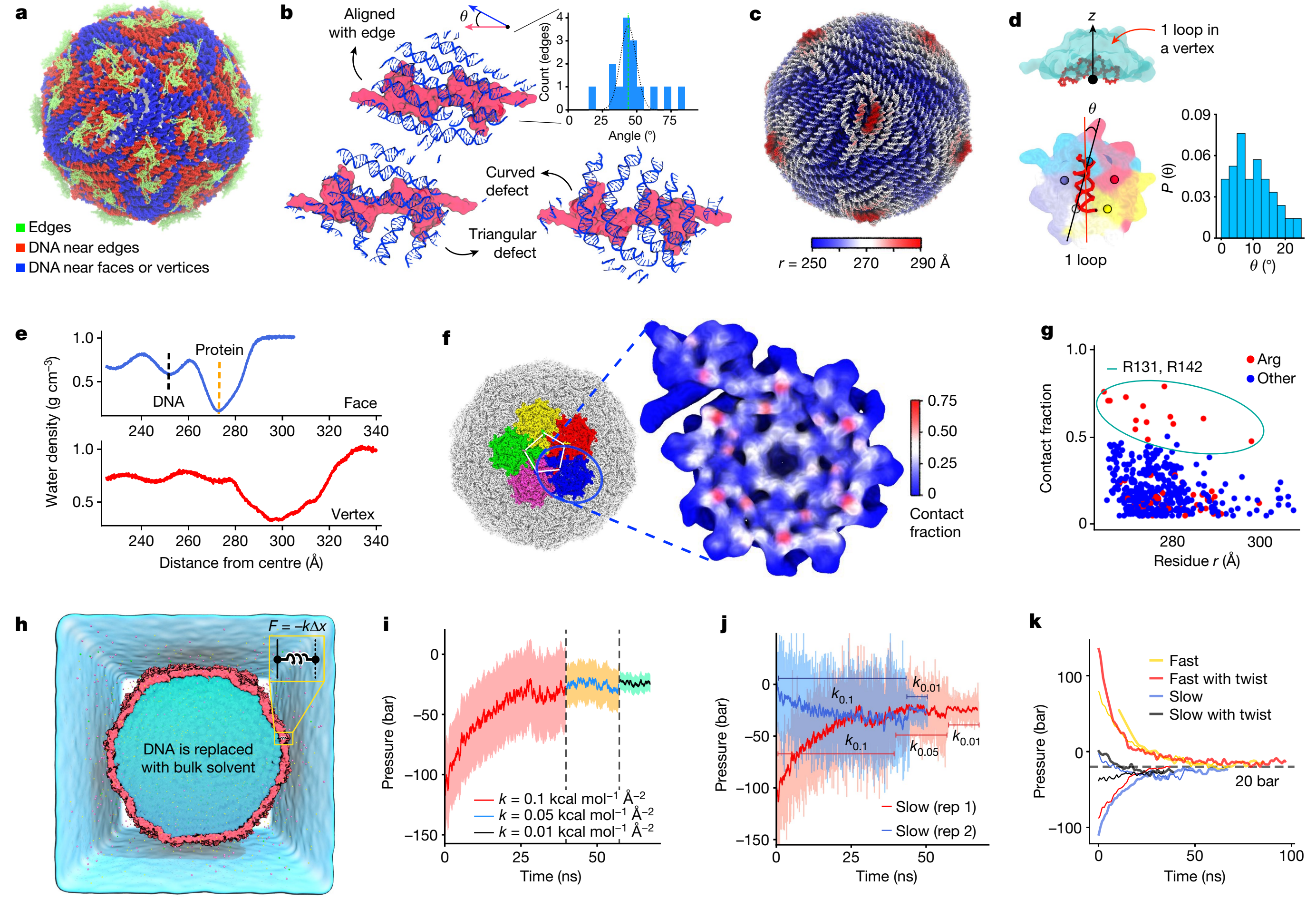


Fig. 5 | Protein–DNA interactions within a packaged capsid. a, Outer layer of the DNA genome coloured in red near the capsid edges and blue near the capsid faces or vertices. Proteins forming the edge of the capsid are shown in green, all other parts of the packaged assembly are not shown. **b**, Examples of DNA (blue) arrangement near specific capsid edges (pink). The inset shows the distribution of DNA orientation relative to the capsid edges. **c**, The outer layer of the DNA genome coloured by its radial distance from the capsid centre. **d**, Example of a DNA loop protrusion into a capsid vertex (top). Orientation of the DNA loop relative to a vector connecting the CoMs of the two nearest capsid proteins (bottom left) and the distribution of the loop orientation (bottom right). **e**, Water density near the DNA–protein interface averaged along the face (top) and vertex (bottom) symmetry axes. **f**, Trajectory-averaged contact map of the protein heptamer with the DNA. **g**, Fraction of time that heptamer residues make contact with the DNA versus the radial coordinate of the residues. Arginine residues are shown in red, all other residue types in blue. Data for Arg131 and Arg142 are circled in teal. **h**, Schematic of the pressure calculation protocol. **i**, Simulated internal pressure for several values of the spring constant. Raw data (faded background) were sampled every 9.6 ps and running averaged with a 0.96 ns window. **j**, Pressure determination simulations performed starting from two instances (rep 1 and rep 2) from the same packaged equilibration trajectory. **k**, Pressure determination simulation for four packaged particle models. Thin lines denote replica simulations started from a different instance of the same trajectory. For clarity, only the running average of the pressure data are shown.

was 2 Å larger near the vertices of the capsid. The plots of the radial water density near the capsid faces and vertices (Fig. 5e) indicated the presence of water between the capsid and the outermost layer of the genome. The water density exhibited well-defined undulations along the radial coordinate, anticorrelating with the DNA density. The water density inside the capsid was approximately 70% of the bulk water density, congruent with the high density of the packaged DNA. The protein capsid formed long-lasting contacts with the DNA (Fig. 5f), most prominently with Arg131 and Arg142 (Fig. 5g) at the inner surface of the capsid.

To determine the pressure inside the packaged capsids, we measured the forces required to maintain the expanded configuration of the packaged particle in the absence of the DNA (Fig. 5h and Supplementary Video 10; see Methods for details). Figure 5i illustrates the results of one such restrained simulation. Although the instantaneous pressure exhibited large fluctuations, its average gradually converged to a steady-state value. Moreover, the magnitude of the pressure fluctuation depended on the spring constant of the restraints, whereas the average value did not. To verify convergence of the pressure data, we repeated the pressure determination simulations using, as a restraint target, a configuration taken from the same trajectory 50 ns before the configuration used for the original pressure determination simulation (Fig. 5j). Repeating the two sets of pressure calculation for another three models of the packaged particle provided eight estimates of the internal pressure (Fig. 5k). Within the resolution of our method of pressure determination, we did not find a systematic dependence of the internal pressure on the presence of twist during the packaging process. Averaging over the 8 steady-state values resulted in the internal pressure of the HK97 capsid at 21 ± 2.0 bar, which is within the range of experimental values, 20–60 bar, as measured for other bacteriophages[6,56,57].

Conclusions

We presented a computational approach to obtain complete all-atom models of fully packaged virions enabled by a multiresolution representation of DNA and proteins. Our computational reconstruction of a bacteriophage virion led to several unexpected results. First, our model predicted that genome packaging occurs through a loop extrusion mechanism, which, by form, resembles the loop extrusion process implicated in large-scale genome organization[7]. This raises the question of whether loop extrusion could be a general mechanism for shaping DNA in dense environments. Furthermore, the 3D structures of the packaged genomes varied substantially from one particle to another because of distinct topological defects that emerge midway through the packaging process. It remains to be seen whether such differences in the genome structure could account for the heterogeneity of the genome ejection process[45,58] or, more broadly, affect the infectivity of individual virions.

Multiple all-atom simulations of fully packaged particles provided a wealth of information not easily accessible by other means. We found that the packaged DNA gently pushes away the faces of the capsid, reducing their fluctuations and the rate of water and ion transport across and within the capsid. Electrostatic analysis of the all-atom MD trajectories showed that the packaged particles have a neutral inner core surrounded by a polarized protein shell. The physical properties of the packaged particles were insensitive to the presence of the twist during the packaging process.

As with any structure prediction method, our multiresolution approach has limitations that present opportunities for future development. The action of the packaging motor can be represented more accurately when new experiments uncover additional mechanistic detail. The packaging simulations themselves can be performed even slower, in stages that reproduce the natural maturation process of a virus and using more detailed representation of DNA–capsid interactions. The all-atom simulation methodology can be improved to account for changes in the local electrostatic environment produced by titratable residues and for slow binding kinetics of divalent ions. Finally, packaging DNA with less force into a flexible capsid with explicit ions, a more detailed description of friction and longer all-atom equilibration will further increase the realism of our methodology.

Our multiresolution simulation strategy opens the door to obtaining atomically precise, complete 3D structural models of multiple viral species, potentially at different stages of viral infection. When applied on an exascale supercomputer, our method will be able resolve the 3D structures of large, pressurized viruses, including that of the herpes simplex virus. The multiresolution approach can be extended to other types of viruses when combined with the information on the secondary structure of their genomes and their local affinity to the protein capsids. The all-atom structural models of fully packaged virions offer a possibility of finding reproducible structural features at the genome–capsid interface or within the genome configuration. Indeed, we already saw such features in the outer layer of the HK97 genome, imprinted by the icosahedral confinement of the capsid. The exploration of such features as drug targets may offer physical means to treat viral infections, resistant against mutations in the viral genome sequence.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-07150-4.

1. Jiang, W. & Tang, L. Atomic cryo-EM structures of viruses. *Curr. Opin. Struct. Biol.* **46**, 122–129 (2017).
2. Luque, D. & Castón, J. R. Cryo-electron microscopy for the study of virus assembly. *Nat. Chem. Biol.* **16**, 231–239 (2020).
3. Dai, X. et al. In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature* **541**, 112–116 (2017).
4. Ilca, S. et al. Multiple liquid crystalline geometries of highly compacted nucleic acid in a dsRNA virus. *Nature* **570**, 252–256 (2019).
5. Duda, R. L. & Teschke, C. M. The amazing HK97 fold: versatile results of modest differences. *Curr. Opin. Virol.* **36**, 9–16 (2019).
6. Smith, D. E. et al. The bacte-riophage φ29 portal motor can package DNA against a large internal force. *Nature* **413**, 748 (2001).
7. Ganji, M. et al. Real-time imaging of DNA loop extrusion by condensin. *Science* **360**, 102–105 (2018).
8. Speir, J. A. & Johnson, J. E. Nucleic acid packaging in viruses. *Curr. Opin. Struct. Biol.* **22**, 65–71 (2012).
9. Johnson, J. E. & Chiu, W. DNA packaging and delivery machines in tailed bacteriophages. *Curr. Opin. Struct. Biol.* **17**, 237–43 (2007).
10. Zandi, R., Dragnea, B., Travesset, A. & Podgornik, R. On virus growth and form. *Phys. Rep.* **847**, 1–102 (2020).
11. Tzlil, S., Kindt, J. T., Gelbart, W. M. & Ben-Shaul, A. Forces and pressures in DNA packaging and release from viral capsids. *Biophys. J.* **84**, 1616–1627 (2003).
12. Purohit, P. K., Kondev, J. & Phillips, R. Mechanics of DNA packaging in viruses. *Proc. Natl Acad. Sci. USA* **100**, 3173–8 (2003).
13. Kindt, J., Tzlil, S., Ben-Shaul, A. & Gelbart, W. M. DNA packaging and ejection forces in bacteriophage. *Proc. Natl Acad. Sci. USA* **98**, 13671–13674 (2001).
14. Perlmutter, J. D., Qiao, C. & Hagan, M. F. Viral genome structures are optimal for capsid assembly. *eLife* **2**, e00632 (2013).
15. Huang, R. K. et al. The prohead-I structure of bacteriophage HK97: implications for scaffold-mediated control of particle assembly and maturation. *J. Mol. Biol.* **408**, 541–554 (2011).
16. Earnshaw, W. C. & Harrison, S. C. DNA arrangement in isometric phage heads. *Nature* **268**, 598–602 (1977).
17. Duda, R. L. et al. Structure and energetics of encapsidated DNA in bacteriophage HK97 studied by scanning calorimetry and cryo-electron microscopy. *J. Mol. Biol.* **391**, 471–483 (2009).
18. Lander, G. C. et al. The structure of an infectious P22 virion shows the signal for headful DNA packaging. *Science* **312**, 1791–1795 (2006).
19. Bloomfield, V. A. DNA condensation by multivalent cations. *Biopolymers* **44**, 269–282 (1997).
20. Grosberg, A. Y., Nguyen, T. T. & Shklovskii, B. I. Colloquium: the physics of charge inversion in chemical and biological systems. *Rev. Mod. Phys.* **74**, 329–345 (2002).
21. Fuller, D. N. et al. Ionic effects on viral DNA packaging and portal motor function in bacteriophage φ29. *Proc. Natl Acad. Sci. USA* **104**, 11245–11250 (2007).
22. Evilevitch, A. et al. Effects of salts on internal DNA pressure and mechanical properties of phage capsids. *J. Mol. Biol.* **405**, 18–23 (2011).
23. Huber, R. G., Marzinek, J. K., Holdbrook, D. A. & Bond, P. J. Multiscale molecular dynamics simulation approaches to the structure and dynamics of viruses. *Prog. Biophys. Mol. Biol.* **128**, 121–132 (2017).
24. Casalino, L. et al. AI-driven multiscale simulations illuminate mechanisms of SARS-CoV-2 spike dynamics. *Int. J. High Perform. Comput. Appl.* **35**, 432–451 (2021).
25. Yu, A. et al. Strain and rupture of HIV-1 capsids during uncoating. *Proc. Natl Acad. Sci. USA* **119**, e2117781119 (2022).
26. Cerritelli, M. E. et al. Encapsidated conformation of bacteriophage T7 DNA. *Cell* **91**, 271–280 (1997).
27. Forrey, C. & Muthukumar, M. Langevin dynamics simulations of genome packing in bacteriophage. *Biophys. J.* **91**, 25–41 (2006).
28. Locker, C. R., Fuller, S. D. & Harvey, S. C. DNA organization and thermodynamics during viral packing. *Biophys. J.* **93**, 2861–2869 (2007).
29. Petrov, A. S., Boz, M. B. & Harvey, S. C. The conformation of double-stranded DNA inside bacteriophages depends on capsid size and shape. *J. Struct. Biol.* **160**, 241–248 (2007).
30. Marenduzzo, D. et al. DNA–DNA interactions in bacteriophage capsids are responsible for the observed DNA knotting. *Proc. Natl Acad. Sci. USA* **106**, 22269–22274 (2009).
31. Dans, P. D. et al. The static and dynamic structural heterogeneities of B-DNA: extending Calladine–Dickerson rules. *Nucleic Acids Res.* **47**, 11090–11102 (2019).
32. Liebl K. & Zacharias M. The development of nucleic acids force fields: From an unchallenged past to a competitive future. *Biophys. J.* https://doi.org/10.1016/j.bpj.2022.12.022 (2023).
33. Yoo, J., Winogradoff, D. & Aksimentiev, A. Molecular dynamics simulations of DNA–DNA and DNA–protein interactions. *Curr. Opin. Struct. Biol.* **64**, 88–96 (2020).
34. Lynch D. L., Pavlova A., Fan Z. & Gumbart J. C. Understanding virus structure and dynamics through molecular simulations. *J. Chem. Theory Comput.* https://doi.org/10.1021/acs.jctc.3c00116 (2023).
35. Perilla, J. R. & Schulten, K. Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations. *Nat. Commun.* **8**, 15959 (2017).
36. Hadden, J. A. et al. All-atom molecular dynamics of the HBV capsid reveals insights into biological function and cryo-EM resolution limits. *eLife* **7**, e32478 (2018).
37. Durrant, J. D. et al. Mesoscale all-atom Influenza virus simulations suggest new substrate binding mechanism. *ACS Cent. Sci.* **6**, 189–196 (2020).
38. Bryer, A. J., Reddy, T., Lyman, E. & Perilla, J. R. Full scale structural, mechanical and dynamical properties of HIV-1 liposomes. *PLoS Comput. Biol.* **18**, e1009781 (2022).
39. Freddolino, P. L., Arkhipov, A. S., Larson, S. B., McPherson, A. & Schulten, K. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure* **14**, 437–449 (2006).
40. Gan, L. et al. Capsid conformational sampling in HK97 maturation visualized by X-ray crystallography and cryo-EM. *Structure* **14**, 1655–1665 (2006).
41. May, E. R., Arora, K. & Brooks, C. L. III pH-induced stability switching of the bacteriophage HK97 maturation pathway. *J. Am. Chem. Soc.* **136**, 3097–3107 (2014).
42. Jana, A. K., Sharawy, M. & May, E. R. Non-equilibrium virus particle dynamics: microsecond MD simulations of the complete flock house virus capsid under different conditions. *J. Struct. Biol.* **215**, 107964 (2023).
43. Andoh, Y. et al. All-atom molecular dynamics calculation study of entire poliovirus empty capsids in solution. *J. Chem. Phys.* **141**, 165101 (2014).
44. Liu, S. et al. A viral packaging motor varies its DNA rotation and step size to preserve subunit coordination as the capsid fills. *Cell* **157**, 702–713 (2014).

45. Grayson, P., Han, L., Winther, T. & Phillips, R. Real-time observations of single bacteriophage lambda DNA ejections in vitro. *Proc. Natl Acad. Sci. USA* **104**, 14652 (2007).
46. Arsuaga, J., Tan, R. K.-Z., Vazquez, M., Sumners, D. W. & Harvey, S. C. Investigation of viral DNA packaging using molecular mechanics models. *Biophys. Chem.* **101-102**, 475–484 (2002).
47. Berndsen, Z. T., Keller, N., Grimes, S., Jardine, P. J. & Smith, D. E. Nonequilibrium dynamics and ultraslow relaxation of confined DNA during viral packaging. *Proc. Natl Acad. Sci. USA* **111**, 8345–8350 (2014).
48. Vroege, G. J. & Odijk, T. Induced chain rigidity, splay modulus and other properties of nematic polymer liquid crystals. *Macromolecules* **21**, 2848–2858 (1988).
49. Leforestier, A. & Livolant, F. The bacteriophage genome undergoes a succession of intracapsid phase transitions upon DNA ejection. *J. Mol. Biol.* **396**, 384–395 (2010).
50. Yoo, J. & Aksimentiev, A. The structure and intermolecular forces of DNA condensates. *Nucleic Acids Res.* **44**, 2036–2046 (2016).
51. Córdoba, A., Hinckley, D. M., Lequieu, J. & de Pablo, J. J. A molecular view of the dynamics of dsDNA packing inside viral capsids in the presence of ions. *Biophys. J.* **112**, 1302–1315 (2017).
52. Widom, J. & Baldwin, R. L. Tests of spool models for DNA packaging in phage lambda. *J. Mol. Biol.* **171**, 419–437 (1983).
53. Bauer, D. W. et al. Exploring the balance between DNA pressure and capsid stability in herpes and phage. *J. Virol.* **89**, 9288–9298 (2015).
54. Qiu, X. et al. Salt-dependent DNA–DNA spacings in intact bacteriophage λ reflect relative importance of DNA self-repulsion and bending energies. *Phys. Rev. Lett.* **106**, 028102 (2011).
55. Dixit, A., Ray, K., Lakowicz, J. R. & Black, L. W. Dynamics of the T4 bacteriophage DNA packasome motor. *J. Biol. Chem.* **286**, 18878–18889 (2011).
56. Grayson, P. et al. The effect of genome length on ejection forces in bacteriophage lambda. *Virology* **348**, 430–436 (2006).
57. Roos, W., Ivanovska, I., Evilevitch, A. & Wuite, G. Viral capsids: mechanical characteristics, genome packaging and delivery mechanisms. *Cell. Mol. Life Sci.* **64**, 1484–1497 (2007).
58. Chiaruttini, N. et al. Is the in vitro ejection of bacteriophage DNA quasistatic? A bulk to single virus study. *Biophys. J.* **99**, 447–455 (2010).
59. Lu, X. J. & Olson, W. K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.* **3**, 1213–1227 (2008).

Methods

All-atom MD simulations

All atomistic MD simulations were performed using a memory optimized version of NAMD2.13b[60], the CHARMM36 parameter set[61] for protein and DNA, the TIP3P model for water[62] and a custom hexahydrate model for magnesium ions[63] used along with the CUFIX corrections to ion–DNA interactions[63]. A full set of CUFIX corrections[64] was used where specified. The use of permanent magnesium hexahydrate complexes to represent Mg^{2+} ions prevented direct binding of Mg^{2+} to DNA, in line with previous findings[65]. Multiple time stepping[66] was used, whereby local interactions were computed every 2 fs whereas long-range interactions were computed every 6 fs. All short-range nonbonded interactions were cut-off starting at 0.8 nm and completely cut-off by 1.0 nm. Long-range electrostatic interactions were evaluated using the particle-mesh Ewald method[67] computed over a 0.21-nm-spaced grid with an interpolation order of eight. SETTLE[68] and RATTLE[69] algorithms were applied to constrain covalent bonds to hydrogen in water and in non-water molecules, respectively. The temperature was maintained at 300 K using a Langevin thermostat with a damping constant of 0.5 ps^{-1} unless specified otherwise. Constant pressure (NPT ensemble) simulations used a Nosé–Hoover Langevin piston[70] with a period and decay of 1,400 and 700 fs, respectively. Implicit representation of the capsid was realized using the Grid-Steered MD[71] feature of NAMD. Energy minimization was carried out using the conjugate gradients method[72]. Atomic coordinates were recorded every 9.6 ps unless specified otherwise. Visualization and analysis were performed using VMD[73] and MDanalysis[74], respectively. Unless specified otherwise, custom scripts for Tcl interpreter of VMD were used in analysis of trajectories. Supplementary Table 1 provides a summary of the all-atom MD runs performed.

Tiling electrolyte volume for solvation. A cubic volume of electrolyte about 3 nm on one side was first simulated for 30 ns in the NPT ensemble. The last frame of the trajectory was used to build a much larger volume of the electrolyte by copying and shifting the electrolyte coordinates by a predetermined number of unit cell vectors. Because Mg^{2+} is modelled as a hexahydrate complex, water molecules of the hexahydrates were wrapped across the periodic boundaries of the unit cell before tiling. A typical solvation process included merging the tiled electrolyte volume with the protein or DNA structure followed by removal of water and ions that overlapped with protein or DNA. Having a slight excess of ions in the tiled volume was instrumental to obtaining a target ion density after solvation.

Vacuum simulations. Vacuum simulations were performed to resolve steric clashes between atoms. Unless stated otherwise, such simulations were performed using a dielectric constant of 200, the long-range electrostatic calculations (PME) switched off and a Langevin thermostat with a damping coefficient of 50 ps^{-1}. To ensure the structural integrity of the molecules of interest, non-hydrogen atoms of the molecules were subject to a network of harmonic restraints and/or, in the case of DNA, external bonds enforcing base pairing (see respective sections for the specific values of the parameters). Under such restraints, the overall RMSD of the molecule was about 0.5 Å.

All-atom simulations of empty HK97 capsid

An initial all-atom model of the HK97 capsid was constructed from the structure of the asymmetric subunit[40] obtained by X-ray crystallography and cryo-EM (PDB identifier: 2FT1). First, the amino terminus of each of the seven proteins in the asymmetric subunit was extended to include residues that were not resolved in the structure using a model of the immature asymmetric subunit (PDB identifier: 1OHG)[75]. Note that the N-terminal disordered tails of the capsid proteins (up to residue 103) are cleaved during the maturation process and were, accordingly, omitted from the model. The final protein model included residues 103–383 for all capsid proteins. With crystallographically resolved atoms held fixed, steric clashes were resolved by 1,200 steps of minimization. Subsequently, a model of the complete HK97 capsid was generated by tiling the asymmetric subunit about the symmetry axes of the capsid. The titratable residues located on the inner surface of the capsid, that is, Asp136, Glu149, Asp173, Glu269, His283, Asp290, Glu292, Asp330 and Glu348, were protonated[41], which reduced the electrical charge of the protein capsid to zero. Finally, Lys169 and Asn356 residues of adjacent proteins were crosslinked using a custom patch (Supplementary Information). Parameters for the atoms affected by the crosslinking were obtained using the CGenFF server[76].

The capsid was merged with a 64 × 64 × 64 nm^3 tiled volume of solvent with an excess concentration of ions. Water within 1.24 Å and ions within 2.25 Å of non-hydrogen atoms of the capsid were removed from the system. Ions were then removed from inside and outside the capsid until the Na^+, Mg^{2+} and Cl^- concentrations were, respectively, approximately 0.21, 0.005 and 0.3 mol kg^{-1} inside and approximately 0.22, 0.028 and 0.24 mol kg^{-1} outside the capsid. At the end of the solvation procedure, the system was electrically neutral and contained 27,484,866 atoms. Following 9,600 steps of energy minimization, the solvated system was equilibrated in multiple stages (Supplementary Table 2) of NPT simulations at 300 K. With non-hydrogen atoms of the capsid subject to position restraints ($k = 0.1$ kcal mol^{-1} $Å^{-2}$), a gradual lowering of the Langevin damping coefficient from 50 ps^{-1} (for 0.5 ns) to 10 ps^{-1} (for 2.2 ns) and finally to 0.5 ps^{-1} (for 18.8 ns) allowed water and ions to equilibrate, bringing the volume of the system to equilibrium in about 10 ns. Next, position restraints on the protein were gradually decreased in stages. First, the system was equilibrated for another 17.14 ns with the position restraints applied to all non-hydrogen atoms of the capsid ($k = 0.01$ kcal $mol^{-1}Å^{-2}$). Then the restraints were lowered ($k = 0.005$ kcal mol^{-1} $Å^{-2}$) and simulated for 14.27 ns, after which, only the Cα atoms of the protein were restrained ($k = 0.005$ kcal mol^{-1} $Å^{-2}$) for 12.01 ns. As a final step, the position restraints were further lowered ($k = 0.001$ kcal $mol^{-1}Å^{-2}$) for 4.7 ns, after which, the system was run without any position restraints. The last 600 ns of the simulation was performed using CUFIX corrections applied to acetate–ion and amine–carboxyl interactions. A replica simulation of the empty capsid was built and equilibrated following the same procedures, differing by the initial concentration of ions, 0.2 mol kg^{-1} Na^+, 0.005 mol kg^{-1} Mg^{2+} and 0.21 mol kg^{-1} Cl^-, both inside and outside the capsid. The equilibration schedule is further detailed in Supplementary Tables 2 and 3.

Multiresolution simulations of DNA packaging

DNA packaging simulations were performed using a previously described multiresolution model of DNA, mrDNA[77]. The mrDNA model was developed to reproduce experimentally determined properties of DNA such as its persistence length and twist persistence length, as well as the pressure in a DNA array. Packaging simulations were performed using an in-house-developed, GPU-accelerated code, atomic resolution Brownian dynamics (ARBD)[78]. All simulations used a Brownian dynamics integrator with simulation-dependent time step, a temperature of 291 K and a nonbonded cut-off value of 4.1 nm or 5 nm for the fast or slow packaging simulations, respectively. Although the smoothening of the free-energy landscape makes processes occur much faster in a CG simulation than in an equivalent all-atom simulation[79], the durations of our CG simulations are reported throughout the article simply as the product of the time step and the number of elapsed steps.

Grid-based representation of the capsid enclosure and the packaging motor. For our DNA packaging simulations, the protein capsid containing an embedded packaging motor was represented using a 3D grid-based potential. Before generating such a potential, a small patch of atoms was cut out from the all-atom structure of the capsid (PDB identifier: 2FT1)[40] at one of the icosahedral vertices of the capsid. An all-atom model of an HK97-family portal protein (PDB identifier: 3KDR) was manually docked into the vertex opening. A 2.06 Å-resolution steric potential representing the capsid and the embedded portal was generated from the all-atom model using the volmap tool of VMD[73], with each non-hydrogen atom replaced by a spherical Gaussian density of width equal to the nominal atomic radius multiplied by a scaling factor of 2. To model the action of the packaging motor, a linear potential ramp was added to the cylindrical region (15 Å in radius and 63.7 Å in length) inside the portal, as illustrated by the inset in Fig. 2a. To avoid entrapment of DNA beads within local wells of the capsid potential, a spherical Gaussian kernel of 1 Å width was linearly convolved with the density of the capsid, blurring the density and ensuring that the derived potential acting on DNA beads is smooth and continuous. This density was multiplied by a scaling factor of 0.5 kcal mol^{-1} to obtain the implicit potential for the CG packaging simulations. To the best of our knowledge, there is no experimental evidence suggesting that specific residues of the HK97 capsid have a crucial role during the packaging process. Experiments performed using a similar virus, phage-λ, found that packaged DNA forms random contacts with the capsid[52], which suggests that representing the capsid as a steric, chemically inert barrier is a reasonable approximation.

Initial configurations of DNA genomes. Before packaging, initial configurations of the 39,732 bp HK97 genome were prepared as follows. The DNA was initially arranged in the shape of a cylindrical 3D spiral that did not contain knots. For fast packaging simulations, the spiral was 108 nm in length and 85 nm in diameter, and the DNA was equilibrated for 5 µs (100 fs time step) in the absence of the capsid potential at 4 bp per 1 bead resolution. The resulting configuration was placed with one end 28 nm away from the capsid portal having the portal and spiral axes coinciding. The terminal DNA bead nearest to the capsid was then pulled into the portal by a harmonic potential, (k = 0.1 kcal mol^{-1} Å^{-2}) during a simulation lasting approximately 0.5 µs. For slow packaging simulations, the DNA genome was initially placed in the shape of a spiral filling a 254-nm-long annular cylinder with 50 nm and 100 nm inner and outer radii, respectively. Four replicate simulations were then performed at 10 bp per bead resolution to relax the configuration, each lasting 2.5 ms (250 fs time step). Each DNA configuration was then geometrically transformed to orient the DNA fragment near one of its ends along the portal axis and place that end 11.5 nm away from the portal entrance. Those configurations were then used to construct a 4 bp per 2 bead model using the mrDNA protocol[77]. The conformations of the newly added orientation beads were relaxed during a 40 ns simulation (40 fs time step), whereas the DNA contour beads were each harmonically restrained to their initial coordinates (k = 10 kcal mol^{-1} Å^{-2}). The DNA end nearest to the portal was then brought to the portal entrance during another 40 ns simulation using a harmonic restraint potential (k = 1 kcal mol^{-1} Å^{-2}), one such configuration is shown in the inset in Fig. 2a. Four additional 4 bp per 2 bead relaxation simulations were performed starting from the mirror-image configurations of the states obtained at the end of the four 10 bp per bead relaxation simulations, which provided a total of eight initial conditions for the slow DNA packaging simulations.

DNA packaging simulations. The genome packaging simulations were performed at either 4 bp per 2 bead resolution with 2 magnitudes of the packaging force, 280 pN (fast) or 55 pN (slow). In some simulations, additional torque was applied to the beads located within the portal region to rotate the DNA right-handed with respect to the axis directed from the portal to the capsid centre by either −359° (fast with twist) or 14° (slow with twist) per turn of the packaged DNA.

The simulations of DNA packaging with twist were performed using a custom build of ARBD that placed a harmonic potential on the angle formed by the vector connecting each DNA bead in the portal region and its orientation bead, projected into the plane defined by the portal axis, and a reference vector. The reference vector was given by a rotation of a vector within a plane normal to the portal axis by an angle θ_0 that depended on the identity of the DNA bead and on whether the genome was packaged fast or slow. For fast packaging simulations, the spring constant was 500 kcal mol^{-1} radian^{-2} and the rest angle was specified (in degrees) as $\theta_0 = (i_{\mathrm{bp}}/10.4) - (d/34) \times 14$, where i_{bp} is the index of the base pair corresponding to the DNA bead, and d is the position of the bead projected along the portal axis, given in Å. For slow packaging simulations, the spring constant was 10 kcal mol^{-1} radian^{-2} and the rest angle was specified (in degrees) as $\theta_0 = i_{\mathrm{bp}}\,(360/10.4 + 14/10.0)$. The first term accounts for the natural twist of the DNA, whereas the second term effectively rotates the DNA by 14° every 10 bp (one turn of a DNA duplex). The 14°-per-turn rotation of DNA was previously reported for the ϕ29 portal[44].

The fast packaging simulations were performed using a 100 fs time step and lasted around 100 µs, whereas the slow packaging simulations were performed with a smaller 40 fs time step and lasted about 1.1 ms. Completion of the slow packaging simulations required about 3 months of continuous runs using the GPU nodes of the TACC Frontera. After completion of the packaging simulations, the systems were equilibrated for another 200 µs, at the same resolution as packaging, that is, 4 bp per 2 beads. This additional equilibration was essential for the fast packaging simulations with high levels of twist, as it allowed the superhelical twist accumulated from the packaging step to equilibrate. In total, 16 slow packaging simulations were performed, 8 each with and without twist; and 4 fast packaging simulations, 2 each with and without twist.

All-atom model of packaged HK97

All-atom model of the genome. After packaging and equilibration of select capsids, the multiresolution protocol[77] was used to map the coordinates of the DNA to a higher resolution, 2 beads per 1 bp, model. The higher-resolution model was equilibrated for 400 µs (slow) or 1 µs (fast), using a smaller time step of 20 fs. The fast model was simulated for another 1 µs at 1 bead per 1 bp resolution. Coordinates from the last part of the final CG trajectory were averaged for mapping to an atomistic model as specified in the mrDNA protocol[77]. First, the trajectory was aligned to minimize the RMSD with respect to the final configuration. Stepping backwards through the trajectory, we selected a continuous fragment of the trajectory with individual DNA configurations and an RMSD with respect to the final configuration lower than a 1 Å threshold. The coordinates from that point onwards were averaged, and splines were fit through these to facilitate mapping to an all-atom structure of the DNA. The nucleotide sequence of the resulting all-atom DNA model matched the experimentally determined sequence of the 39,732 bp HK97 genome[80]. For accurate mapping of atomic coordinates and to prevent any ring piercing clashes from the mapping procedure, the sugar base of the DNA backbone was shrunk in size by a scaling factor of 0.25.

Implicit capsid potential for all-atom simulations. For atomistic simulations, an implicit representation of the capsid was designed to confine DNA while permitting the exchange of water and ions. To obtain the optimal grid potential using the volmap tool of VMD, with the '-weight mass' option turned on, the resolution of the grid and the width of the Gaussians approximating atomic densities were tuned. The resulting grid-based model of the capsid was tested in a vacuum simulation in which an all-atom model of a two-turn DNA fragment was subject to a 20 pN force pressing the DNA duplex onto a small patch of the capsid grid (Extended Data Fig. 10). The specific value of the force was chosen to match that produced by the internal pressure of 20 bar inside a bacteriophage virus. For reference, an identical simulation was performed using a fully atomistic representation of both the DNA and the capsid. The parameters of the grid representations of the capsid were varied until a match with all-atom representation was reached with respect to the equilibrium DNA–capsid distance.

See the caption of Extended Data Fig. 10 for more detail. The resulting grid-based density of the capsid had a resolution of 1.4 Å, and the atomic densities were approximated using Gaussians of the nominal atomic radii. For subsequent grid-steered MD simulations, the grid-based density was converted to a grid-based potential using a scaling factor of 50 kcal mol^{-1} $Å^{-1}$. The instantaneous grid-based forces were computed using a cubic polynomial interpolation in the three dimensions.

Initial relaxation of the genome. A short 2,500-step minimization with backbone atoms fixed was performed in vacuum to bring each DNA nucleotide to its nominal size. Subsequently, another minimization of 4,500 steps in vacuum was performed without having any atoms fixed, but under an elastic network of restraints applied to DNA, as previously described[81]. At the same time, the grid-forces functionality of NAMD[71] was used to apply external forces from the implicit representation of the capsid to non-hydrogen atoms of the DNA, which prevented the DNA from expanding beyond the boundaries of the capsid. Using the extrabonds feature of NAMD, additional harmonic forces (k = 10 kcal mol^{-1} $Å^{-2}$) were applied to pairs of atoms that formed Watson–Crick hydrogen bonds between all base pairs of the genome. The DNA was simulated in vacuum for 21 ns in the NVT ensemble, producing an initial all-atom configuration of the 39,732 bp genome consisting of 2,523,065 atoms. In total, six such all-atom structures were constructed, three each for packaging with and without twist.

Solvation of the genome. To solvate the genome, we prepared a 64 × 64 × 64 nm^3 tiled volume of solvent containing 1 mol kg^{-1} NaCl and 1.2 mol kg^{-1} Mg^{2+} hexahydrate; the solvent carried a net positive charge. This ion composition was chosen to exceed the equilibrium concentration of ions in a tight DNA bundle of approximately the same average DNA–DNA spacing[50]. After addition of the solvent, water and ions overlapping with the atoms of DNA were removed, setting the target concentration for individual ion species by tuning the cut-off distance for the ion removal. For the slow model of packaged DNA, the desired ion composition of 0.75 mol kg^{-1} Na^+ and 0.5 mol kg^{-1} Mg^{2+} (ref. 50) was obtained by removing water molecules with oxygen atoms within 2.3 Å of any non-hydrogen DNA atom, Na^+ ions within 3.7 Å from the DNA, magnesium hexahydrate complexes with Mg^{2+} atoms within 6.855 Å of DNA and Cl^- ions within 20 Å of the DNA, which ensured that no Cl^- ions were initially present inside the volume occupied by packaged DNA. For other DNA models, the above cut-off parameters were slightly adjusted (by less than 5%) to obtain the desired ion concentrations. The ions located outside the volume occupied by the genome were randomly removed such that the final concentrations of ionic species outside the capsid were 0.2 mol kg^{-1} Na^+, 0.005 mol kg^{-1} Mg^{2+} and 0.21 mol kg^{-1} Cl^-. The regions interior of and exterior to the capsid were identified using the 'measure volinterior' functionality of VMD[82]. The resulting solvated systems contained about 26 million atoms.

Equilibration of the solvated genome. The solvated genome structure was equilibrated in a series of simulations described in detail for all systems in Supplementary Tables 4–9. Below, we illustrate the procedure using the slow model as an example. After 2,500 steps of energy minimization, the system was simulated in an NPT ensemble for 1.64 ns with all non-hydrogen atoms of the DNA harmonically restrained (k = 0.1 kcal mol^{-1} $Å^{-2}$) to their initial coordinates and coupling each non-hydrogen atom to the Langevin thermostat with a damping coefficient of 50 ps^{-1}. During this and subsequent simulations, a grid-based potential derived from the atomic structure of the capsid was applied on all non-hydrogen atoms of the DNA to confine it within the inner volume of the capsid, whereas water and ions were allowed to move in and out of the volume occupied by the packaged DNA. Doing so greatly reduced the possibility of having an out-of-equilibrium Mg^{2+} configuration within the genome that could be caused by the low permeability of the protein capsid to Mg^{2+}. After a brief 0.25 ns simulation with the Langevin damping coefficient decreased to 10 ps^{-1}, and another 8.37 ns with the damping coefficient decreased to 0.5 ps^{-1}, the system was simulated for 3.8 ns with the spring constant of the position restraints lowered to 0.01 kcal mol^{-1} $Å^{-2}$. The position restraints were then removed and the system was simulated for 29.13 ns in the NVT ensemble under the action of the confining potential. At all stages of the solvent equilibration procedure, the integrity of all Watson–Crick base pairs of the DNA was enforced by means of external harmonic potentials (k = 50 kcal mol^{-1} $Å^{-2}$).

Addition of the atomistic protein capsid. To replace the implicit representation of the capsid with a corresponding all-atom model, ions residing within the volume to be occupied by the capsid were removed during a 2 ns NVT simulation. During this simulation, all non-hydrogen atoms of the DNA were harmonically restrained (k = 0.1 kcal mol^{-1} $Å^{-2}$) to their coordinates from the end of the previous implicit capsid run, and a grid-based potential was applied only to ions to expel them from the volume to be occupied by the capsid. The ion-expelling grid-based potential was constructed by first creating a 2 Å resolution atomic density of the protein capsid, using the volmap tool in VMD with a scaling factor of 2, and then rescaling the grid cell size to 1.97 Å. This led to the density corresponding to the protein shrinking radially inwards by about 4.5 Å. The resulting density grid was scaled by a factor of 100 kcal mol^{-1} Da^{-1} to produce an external potential. A very small number of ions that, despite the expelling potential, remained in the volume to be occupied by the explicit capsid were manually moved to the region outside the capsid. At the end of this ion-expelling procedure, the volume to be occupied by the capsid was devoid of ions, which necessitated only the removal of water molecules in that area before the addition of the atomistic capsid, which was done as described below.

Next, the atomic structure of the genome was extracted from the last frame of the simulation described in the previous section and merged with the all-atom model of the capsid. A small number of DNA–protein clashes was resolved in a 1 ns vacuum simulation in which all non-hydrogen atoms of the DNA located within 15 Å of the protein were harmonically restrained (k = 0.1 kcal mol^{-1} $Å^{-2}$) to their initial coordinates, whereas the coordinates of all other non-hydrogen atoms of the DNA were constrained to their initial positions using the fix atoms protocol of NAMD. At the same time, the protein capsid was subject to an elastic network of restraints (k = 20 kcal mol^{-1} $Å^{-2}$) applied to all atom pairs located within 6 Å of each other. This vacuum equilibration simulation was performed using a Langevin damping coefficient of 500 ps^{-1} applied to all non-hydrogen atoms. The simulation successfully removed all DNA–protein clashes, perturbing coordinates of the protein atoms by less than 0.7 Å, with the average protein RMSD with respect to the crystal structure coordinates being just 0.2 Å. The resulting DNA–protein system was merged with the solvent configuration obtained at the end of the ion-expelling simulation. Water molecules located within a 2.5 Å of non-hydrogen atoms of the protein capsid were removed. Steric clashes between the protein and the ions were identified, and the ion coordinates were adjusted to remove such clashes, making sure the ion concentration inside and outside the capsid was not disturbed. A small number of persistent clashes with Mg^{2+} hexahydrates were removed by shrinking the size of the hexahydrates by 0.75 followed by a 2,500-step minimization. A similar procedure was performed to build initial models of all packaged particles. Details are described in Supplementary Tables 4–9.

Equilibration of the complete all-atom models. The DNA–protein–solvent systems were next equilibrated in the NPT ensemble as detailed in Supplementary Tables 4–9. Below, we illustrate the equilibration process using the slow model of packaged DNA as an example. The initial equilibration was carried out with position restraints (k = 0.1 kcal mol^{-1} $Å^{-2}$) applied to all non-hydrogen atoms of the protein and DNA, extra bonds applied to preserve the DNA base pairing

($k = 10$ kcal mol^{-1} Å^{-2}) and coupling all non-hydrogen atoms of the system to a Langevin thermostat with a damping coefficient of 50 ps^{-1} for the initial 3.17 ns, 10 ps^{-1} for the next 0.4 ns and 0.5 ps^{-1} for the remaining 21.75 ns. Following that, the position restraints on the DNA were removed and the system was simulated for 14.88 ns while keeping the restraints enforcing the DNA base pairing and the position restraints on the capsid. Then, the position restraints on the non-hydrogen atoms of the protein were lowered to 0.01 kcal mol^{-1} Å^{-2} for 13.53 ns and then to 0.001 kcal mol^{-1} Å^{-2} for another 12.69 ns. Next, position restraints ($k = 0.001$ kcal mol^{-1} Å^{-2}) were applied only to the backbone atoms of the protein for 8.19 ns and then only to the Cα atoms of the protein ($k = 0.001$ kcal mol^{-1} Å^{-2}) for 8.57 ns. The spring constants of the latter restraints were then lowered to 0.0001 kcal mol^{-1} Å^{-2} for the next 11.8 ns along while the spring constant of the DNA base pairing restraints was reduced to 1 kcal mol^{-1} Å^{-2}. For the next 22.72 ns, the system was simulated without any position restraints applied to the protein, whereas the strength of the DNA base pairing restraints was reduced to 0.1 kcal mol^{-1} Å^{-2}. Following that, the system was simulated without applying any restraints to either protein or DNA for 886.27 ns. These steps are described in Supplementary Table 4.

Other all-atom models of packaged capsid were constructed and equilibrated using similar procedures (see Supplementary Tables 5–9 for details). Some of the fast models were equilibrated without applying the NBFIX corrections to acetate–cation and amine–carboxyl interactions[83], whereas the NBFIX corrections for the interactions involving DNA phosphates[63] were present in all simulations. The application of these additional NBFIX corrections slightly increased the gaps between protein interfaces and slightly increased the rate of water transport through the capsid (Fig. 3h).

Common procedures for structure analysis

Unless specified otherwise, the 'measure fit' command of VMD was used for alignment of structures before RMSD calculations, using coordinates of atoms resolved in the crystal structure. Analyses that required local alignment of smaller subunits, for example, per residue RMSD calculations (Fig. 3e), were performed by first aligning all 12 pentameric subunits (one is shown in Fig. 1h) to a chosen reference pentamer using VMD measure fit functionality. After the alignment, the RMSDs of individual residues were computed with respect to the crystallographic coordinates. The RMSDs were then averaged individually for each of the seven protein locations in the asymmetric subunit Fig. 3e, with each type appearing five times in a pentameric subunit.

Capsid deformation calculations (Figs. 1b and Fig. 3b) were performed on a supercomputer using a Python-based framework, whereby the calculations for every capsid symmetry transformation were distributed over multiple nodes and the computations per transformation were multi-threaded across multiple CPU cores on the respective node. MDAnalysis[74] was used to first load the trajectories one frame at a time, whereafter a global alignment of the capsid was performed using the Newton–Raphson quaternion and an adjoint matrix method[84]. This workflow was validated against the results of the VMD measure fit command. The atomic density profiles were then computed by averaging over 4 × 4 nm^2 centred sections connecting the ten pairs of opposite faces or seven pairs of opposite vertices.

To characterize steady-state RMSF of the capsid proteins (Figs. 1e and Fig. 3d), the average CoM coordinates of each of the 420 capsid proteins were calculated over the last 50 ns of the MD trajectory. For each protein, the RMSF was then measured over that time interval relative to the average CoM coordinates, sampling instantaneous configurations every 48 ps. The CoM coordinates of a protein were calculated using the coordinates of all Cα atoms resolved in the crystal structure. The RMSF values were then averaged over 60 protein copies for each protein location within the asymmetric subunit. Before displacement calculation (Fig. 3f), protein coordinates were averaged over the last 40 ns of the corresponding MD trajectories (slow and empty), sampled every 48 ps. The displacement was computed as per-residue RMSD of the two average structures.

Radial plots of average ion concentration (Figs. 1l and Fig. 3g) or DNA density (Fig. 4c) were obtained by averaging ten 4 × 4 nm^2 centred sections connecting the ten pairs of the twenty opposite faces of the capsid (face symmetry axis) or six pairs of the twelve opposite vertices of the capsid (vertex symmetry axis) and over the last 48 ns of the respective MD trajectory.

All SAXS profiles were calculated using Crysol[85]. Base-pair-level characterization of the packaged genome structures was carried out using Curves+ (ref. 86) and the standard reference frame for the description of nucleic acid geometry[87].

Cryo-EM-like and TEM-like images of the simulated structures

The last 500 ns of the slow all-atom MD trajectory was sampled every 50 ns, rotating the genome and the protein capsid according to each of the 60 icosahedral symmetries. The volmap plugin of VMD was used to produce a volumetric grid of the DNA and protein atomic density with a resolution of 2.5 Å, averaging over both simulation frames and the symmetry axes. The volmap plugin was similarly used to produce 2.5 Å resolution maps of the CG DNA mass density of each CG capsid genome, averaging over the equilibration portion of each trajectory with a stride of 40 ns. Maps were produced with and without averaging over the symmetry axes of the system. The 2D images in Fig. 4a,b were produced by averaging a 6-nm-thick slab of the density passing through the centre of the capsid, then linearly mapping the density from white to black (4–7 Da Å^{-3}). For CG images, the repulsive regions of the capsid potential are depicted using a semi-transparent blue overlay.

Individual TEM-like images (Extended Data Fig. 6b) were obtained for CG trajectories from the maps without symmetry, averaging by selecting a normal vector, finding the coordinate of each 3D voxel in the plane corresponding to the normal vector and binning the voxels by those normal coordinates with 2.5 Å resolution and summing the density in each bin. The images were then produced from the projected density by linearly mapping the summed mass density between 200 and 1,600 Da Å^{-3}, from white to black.

Calculation of diffusion constants

The local diffusion of the DNA genome (Fig. 4i) was characterized using the following distance histogram fitting procedure[88]. First, the CoM of each 10 bp DNA fragment was recorded for the last 288 ns of the equilibration trajectory, sampled every $\Delta t = 0.48$ ns. Next, displacement of each fragment over that time interval was computed, producing a distribution of instantaneous displacements. The latter was fit by the probability distribution describing diffusion:

$$p(r, \Delta t)\, d^3 r = \frac{1}{(4\pi D \Delta t)^{3/2}} e^{\frac{-r^2}{4D\Delta t}} (4\pi r^2 dr),$$

where a particle starting from the origin finds itself in the element of volume defined by r and $r + dr$ after time Δt with the probability $p(r, \Delta t)$. The fit yielded an estimate for the diffusion coefficient, D. To evaluate local diffusion of DNA fragments along specific axes (Extended Data Fig. 5b), the 3D displacements were projected along the respective axes before constructing the corresponding displacement histograms.

The local diffusion constants of water and ions (Fig. 4j and Extended Data Fig. 5c) were computed using the Einstein's relation. For water diffusion, coordinates of the water oxygen atoms were recorded over the last 0.96 ns of the equilibration trajectory, sampled every 9.6 ps. The mean squared displacement values were computed over half the total number of non-overlapping time lags. Each diffusion coefficient was calculated as the slope of a linear fit to the mean squared displacement versus time lag dependence, using the first 0.24 ns of the dependence for the fit. To obtain a radial distance dependence, the average radial coordinate was computed over the first 0.24 ns of each water trajectory.

A radial bin size of about 13.5 Å was large enough to obtain a radial dependence and short enough to get good resolution and statistics. To characterize ion diffusion, ion coordinates were collected over the last 48 ns of the equilibration trajectory, sampled every 9.6 ps. The radial averaging was done using a bin size of 10.4 Å. All other procedures were the same as for the water diffusion calculations.

Volumetric analysis

The measure volinterior[82] functionality of VMD was used to classify all regions of the system as interior, exterior or boundary with respect to the protein capsid. A closed (pore free) quicksurf representation of the protein capsid was generated using VMD by setting the radius scaling to 2.8, the isovalue to 0.5 and the grid spacing to 1.5. The raytracing algorithm used 32 rays for every coordinate frame. The boundary voxels were then eliminated by using the skeletonize algorithm[89] of the scikit-image Python package, ensuring that all parts of the system were classified as interior or exterior.

Electrostatic model of a packaged particle

The model (Extended Data Fig. 2g) assumed that all charge inside the capsid is distributed at the inner surface of 29.5 nm radius sphere, which is enclosed by a 2.5 nm thick spherical shell of dielectric medium representing the protein. Outside the dielectric medium, a Debye–Hückel description of the electrostatics was adopted with a Debye length of 6.39 Å. In addition to the measured charge inside the capsid, an additional charge q_{extra} was placed on the inner surface of the dielectric, with a compensating charge placed on the outer surface. Under this model the potential decays rapidly outside the capsid and the potential inside is uniform. Hence, the potential difference between inside and outside is taken as the potential inside.

Internal pressure calculations

Starting from a chosen frame of an all-atom equilibration simulation of a fully packaged capsid, we created a corresponding all-atom model of an empty capsid by removing the DNA and resolvating the capsid interior with bulk-like solvent (0.2 mol kg^{-1} NaCl and 0.005 mol kg^{-1} $MgCl_2$) using the tiling method. After resolvation, small differences in the protein coordinates (order of a few angstroms) resulted in initial differences in the water density inside and outside the capsid, thereby leading to different values of the pressure at the beginning of these simulations. Following 6,000 steps of minimization, the system was equilibrated restraining all non-hydrogen atoms of the capsid to the coordinates taken from the packaged capsid equilibration (Supplementary Video 10). For each frame of the restrained equilibration, the pressure was calculated by first determining the harmonic force applied on each atom of the capsid from the atom's displacement relative to its restrained coordinate (Fig. 5h). The radial projection of the force on each atom was then divided by the area of a sphere with radius equal to the radial coordinate of the atom. These values were then summed up over all restrained atoms of the capsid to give an instantaneous pressure. The instantaneous pressure values were averaged over the restrained equilibration trajectory. Similar equilibrium pressure values were obtained when only Cα atoms of the capsid were restrained. For each model of the packaged genome, the pressure calculation simulations were performed using two instantaneous configurations taken at the end and 50 ns before the end of the respective equilibration trajectory (Fig. 5j,k).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

CG packaging trajectories and select all-atom MD trajectories with water molecules removed have been deposited in the Illinois Data Bank under accession code IDB-4930709 (ref. 90). The following structures were used from the PDB: 3KDR, 2FT1 and 1OHG. Source data are provided with this paper.

Code availability

The files needed to setup CG packaging simulations and select all-atom MD simulations have been deposited, along with select analysis scripts, in the Illinois Data Bank under accession code IDB-4930709 (ref. 90). Remaining analysis scripts are available upon reasonable request.

60. Phillips, J. C. et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **153**, 044130 (2020).
61. Huang, J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).
62. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
63. Yoo, J. & Aksimentiev, A. Improved parametrization of Li^+, Na^+, K^+, and Mg^{2+} ions for all-atom molecular dynamics simulations of nucleic acid systems. *J. Phys. Chem. Lett.* **3**, 45–50 (2012).
64. 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.* **20**, 8432–8449 (2018).
65. Subirana, J. A. & Soler-Lopez, M. Cations as hydrogen bond donors: a view of electrostatic interactions in DNA. *Annu. Rev. Biophys. Biomol. Struct.* **32**, 27–45 (2003).
66. Batcho, P. F., Case, D. A. & Schlick, T. Optimized particle-mesh Ewald/multiple-time step integration for molecular dynamics simulations. *J. Chem. Phys.* **115**, 4003–4018 (2001).
67. Darden, T. A., York, D. & Pedersen, L. Particle mesh Ewald: an *N*·log(*N*) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–92 (1993).
68. Miyamoto, S. & Kollman, P. A. SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water molecules. *J. Comput. Chem.* **13**, 952–962 (1992).
69. Andersen, H. C. RATTLE: A "velocity" version of the SHAKE algorithm for molecular dynamics calculations. *J. Comput. Phys.* **52**, 24–34 (1983).
70. Martyna, G. J., Tobias, D.J. & Klein, M.L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189 (1994).
71. Wells, D. B., Abramkina, V. & Aksimentiev, A. Exploring transmembrane transport through α-hemolysin with grid-steered molecular dynamics. *J. Chem. Phys.* **127**, 125101 (2007).
72. Payne, M. C., Teter, M. P., Allan, D. C., Arias, T. A. & Joannopoulos, J. D. Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.* **64**, 1045–1097 (1992).
73. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **14**, 33–38 (1996).
74. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).
75. Helgstrand, C. et al. The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 Å resolution. *J. Mol. Biol.* **334**, 885–899 (2003).
76. Vanommeslaeghe, K. & MacKerell, A. D. Jr. Automation of the CHARMM general force field (CGenFF) ii: bond perception and atom typing. *J. Chem. Inf. Model.* **52**, 3155–3168 (2012).
77. Maffeo, C. & Aksimentiev, A. MrDNA: a multi-resolution model for predicting the structure and dynamics of DNA systems. *Nucleic Acids Res.* **48**, 5135–5146 (2020).
78. Comer, J. & Aksimentiev, A. Predicting the DNA sequence dependence of nanopore ion current using atomic-resolution Brownian dynamics. *J. Phys. Chem. C* **116**, 3376–3393 (2012).
79. Zwanzig, R. Diffusion in a rough potential. *Proc. Natl Acad. Sci. USA* **85**, 2029–2030 (1988).
80. Juhala, R. J. et al. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J. Mol. Biol.* **299**, 27–51 (2000).
81. Maffeo, C., Yoo, J. & Aksimentiev, A. De novo prediction of DNA origami structures through atomistic molecular dynamics simulation. *Nucleic Acids Res.* **44**, 3013–3019 (2016).
82. Bryer, A. J., Hadden-Perilla, J. A., Stone, J. E. & Perilla, J. R. High-performance analysis of biomolecular containers to measure small-molecule transport, transbilayer lipid diffusion, and protein cavities. *J. Chem. Inf. Model.* **59**, 4328–4338 (2019).
83. Yoo, J. & Aksimentiev, A. Improved parameterization of amine–carboxylate and amine–phosphate interactions for molecular dynamics simulations using the CHARMM and AMBER force fields. *J. Chem. Theory Comput.* **12**, 430–443 (2016).
84. Liu, P., Agrafiotis, D. K. & Theobald, D. L. Fast determination of the optimal rotational matrix for macromolecular superpositions. *J. Comput. Chem.* **31**, 1561–1563 (2010).
85. Svergun, D., Barberato, C. & Koch, M. H. CRYSOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* **28**, 768–773 (1995).
86. Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D. & Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **37**, 5917–5929 (2009).
87. Olson, W. K. et al. A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* **313**, 229–237 (2001).
88. Anderson, C. M., Georgiou, G. N., Morrison, I., Stevenson, G. & Cherry, R. J. Tracking of cell surface receptors by fluorescence digital imaging microscopy using a charge-coupled device camera. Low-density lipoprotein and influenza virus receptor mobility at 4 °C. *J. Cell Sci.* **101**, 415–425 (1992).
89. Lee, T.-C., Kashyap, R. L. & Chu, C.-N. Building skeleton models via 3-D medial surface axis thinning algorithms. *CVGIP Graph. Models Image Process.* **56**, 462–478 (1994).

90. Coshic K., Maffeo C., Winogradoff D., Aksimentiev A. Select trajectories, simulation setup, and analysis for "The structure and physical properties of a packaged bacteriophage particle". University of Illinois at Urbana-Champaign. https://doi.org/10.13012/B2IDB-4930709_V1 (2024).
91. Aksimentiev, A. & Schulten, K. Imaging α-hemolysin with molecular dynamics: Ionic conductance, osmotic permeability and the electrostatic potential map. *Biophys. J.* **88**, 3745–3761 (2005).

Acknowledgements This work was supported by the National Institute of General Medical Sciences grant R01-GM137015, the National Science Foundation grant PHY-1430124 and the Human Frontier Science Project (RGP0047/2020). ARBD development is supported by the National Science Foundation grant OAC-2311550. The supercomputer time was provided through the Leadership Resource allocation MCB20012 on Frontera of the Texas Advanced Computing Center and the XSEDE allocation MCA05S028. We thank J. Johnson for sharing SAXS data for the empty HK97 capsid, and A. Evilevitch, I. Golding, R. Duda and S. Butcher for discussions.

Author contributions A.A. conceptualized and supervised the work and performed project administration. A.A., K.C. and C.M. designed the methodology. K.C., C.M. and D.W. performed the investigations and visualizations. A.A., K.C., C.M. and D.W. acquired funding. A.A., K.C., C.M. and D.W. wrote the original draft. A.A., K.C. and C.M. reviewed and edited the article.

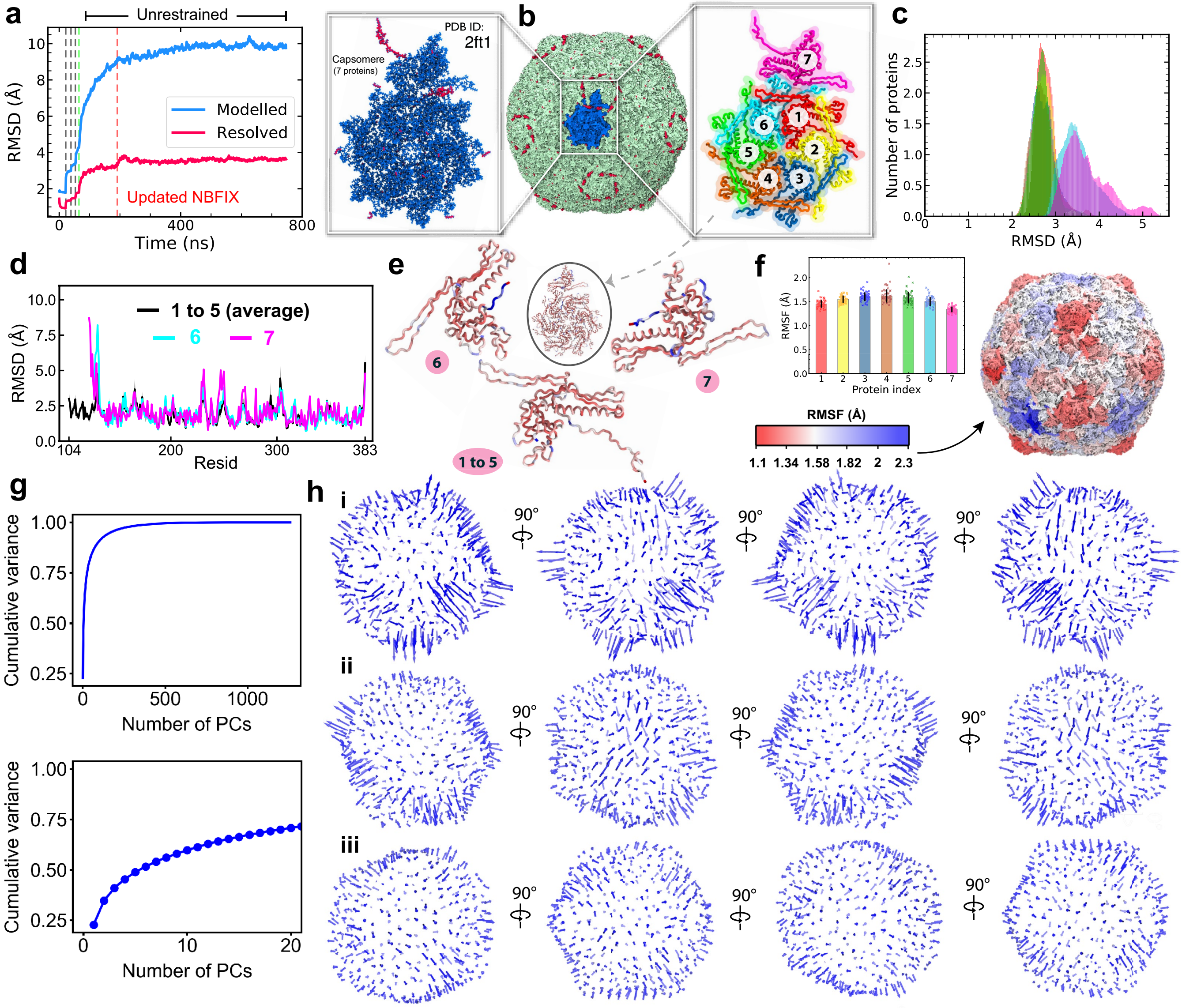
Competing interests The authors declare no competing interests.

Additional information

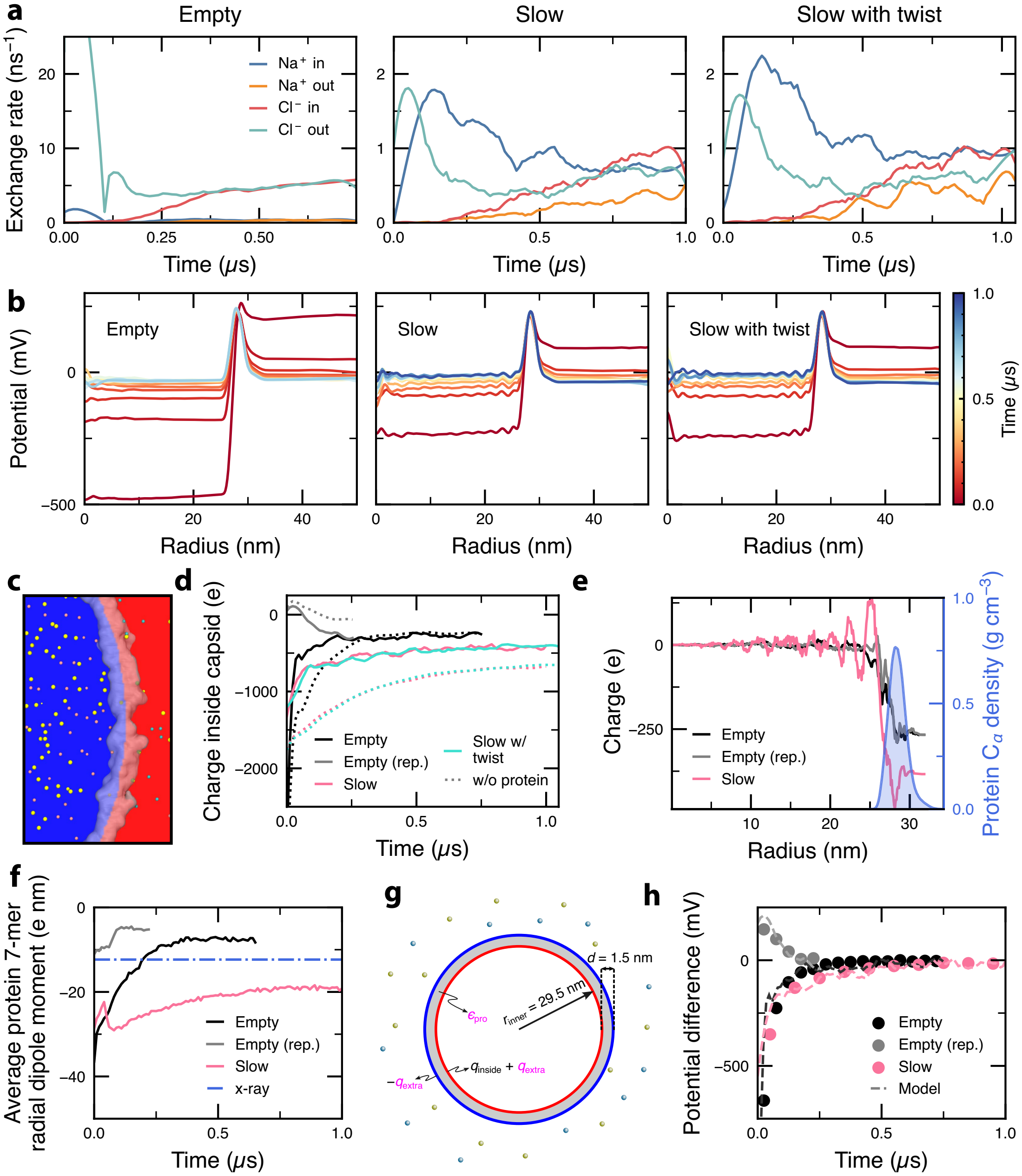
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-07150-4.

Correspondence and requests for materials should be addressed to Aleksei Aksimentiev.

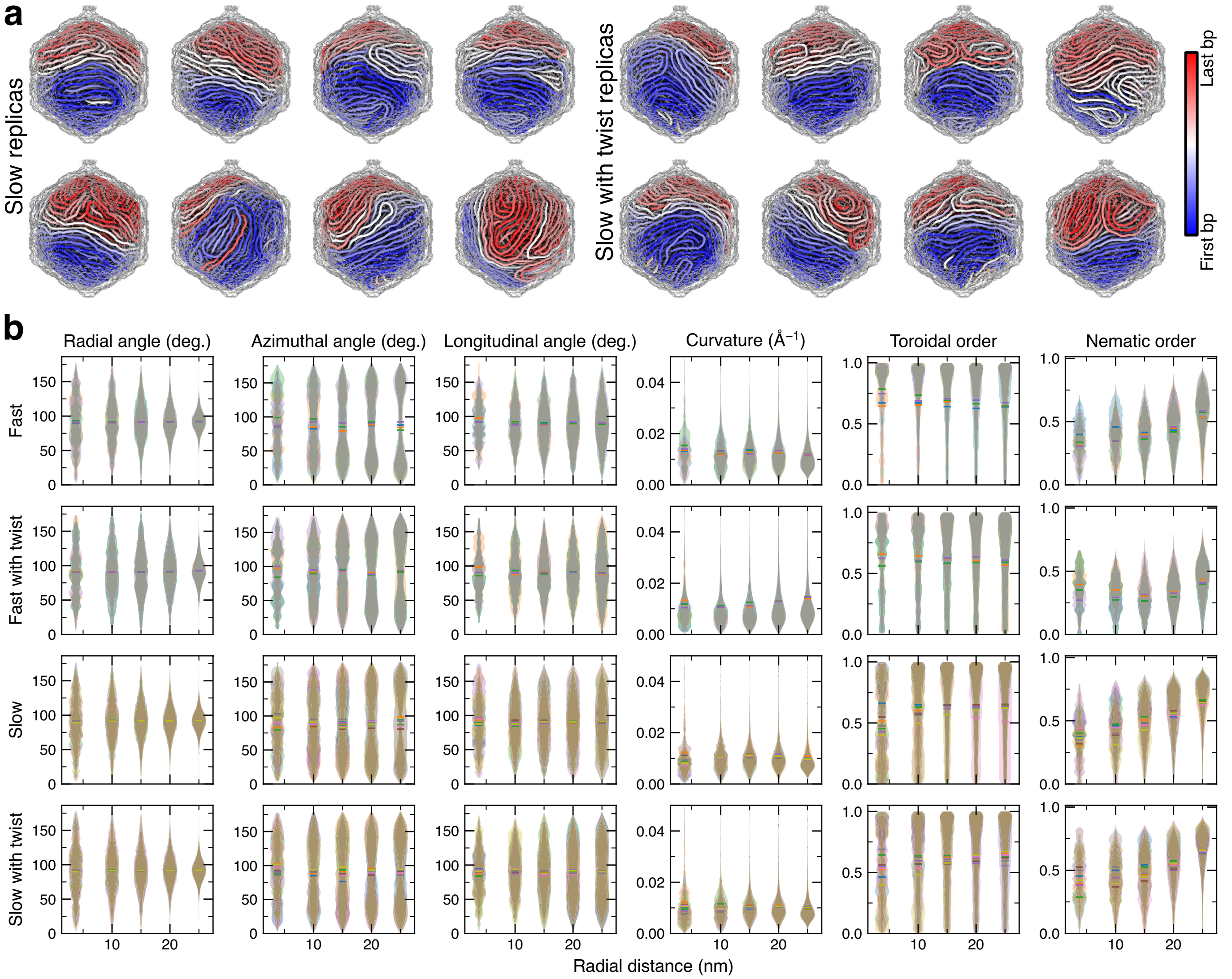
Peer review information *Nature* thanks Eric May and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



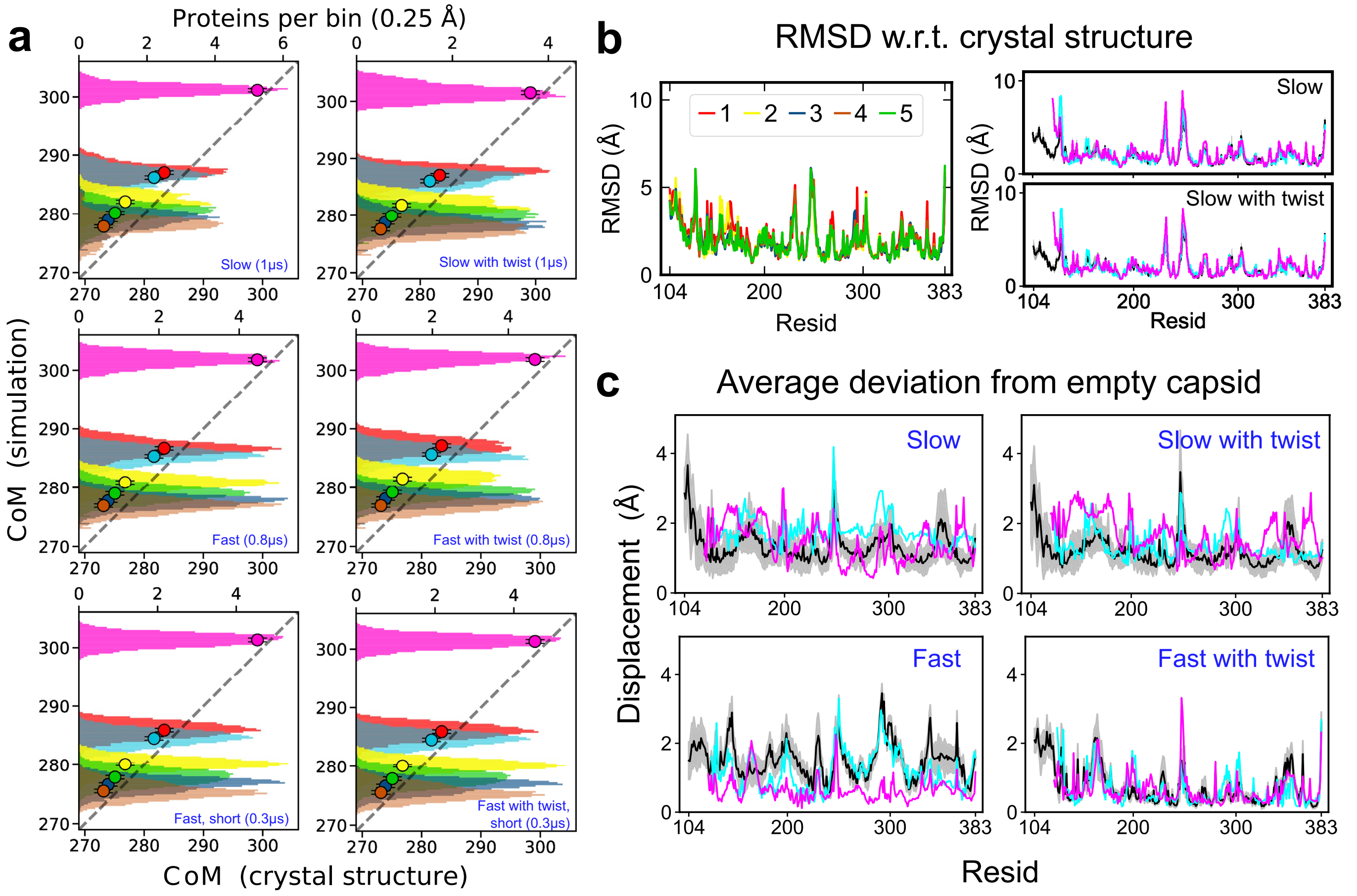
Extended Data Fig. 1 | Structure and fluctuations of empty HK97 capsid. **a**, RMSD of the entire capsid from its initial coordinates during all-atom equilibration simulations. For the first 70 ns, parts of the system were subject to restraints, as detailed in Methods and Supplementary Table 2. The image on the right shows one of the 60 asymmetric subunits with residues resolved in the crystal structure shown in blue and modeled in red. **b**, Arrangement of the asymmetric subunits into an icosahedron capsid (left). Modeled residues are shown red. The right image details the subunit, consisting of seven proteins. **c**, Distribution of individual proteins' average RMSD grouped according to the protein location in the asymmetric subunit. The distributions were computed over the last 38 ns of the equilibration trajectory. Colors are defined in **b**. **d**, Average per-residue RMSD of an empty capsid from the crystallographic coordinates as a function of the residue number. RMSD of modeled residues is not shown. **e**, Cartoon representation of resolved regions of proteins 1 (representative for proteins 1 to 5), 6 and 7 colored according to their average RMSD. **f**, Average RMSF of individual proteins according to their location. The RMSF values were computed over the last 100 ns of the equilibration trajectory. Error bars depict SD over n = 60 copies of the protein. The capsid is shown (right), with individual residues of the proteins colored by their average RMSF. **g**, Cumulative variance of the principal components (PCs). PC analysis was performed using CoM coordinates of each of the 420 capsid proteins and a representative 150 ns fragment of the free equilibration trajectory. (Top) Fractional cumulative variance as a function of the number of PCs, ordered by eigenvalue from highest to lowest. (Bottom) Same, shown only for the top twenty principal components. **h**, Projections of the top three PC, shown as vectors drawn from an average structure. Images in (i), (ii) & (iii) show projections of the first, second and third PCs, ordered by eigenvalue from high to low, each shown from four different perspectives. The top three PCs account for about 23, 12 and 6% of the total variance, respectively. Vector magnitudes drawn are based on the normalized eigenvector multiplied by the corresponding eigenvalue and then multiplied by a factor of ten, to facilitate visualization.



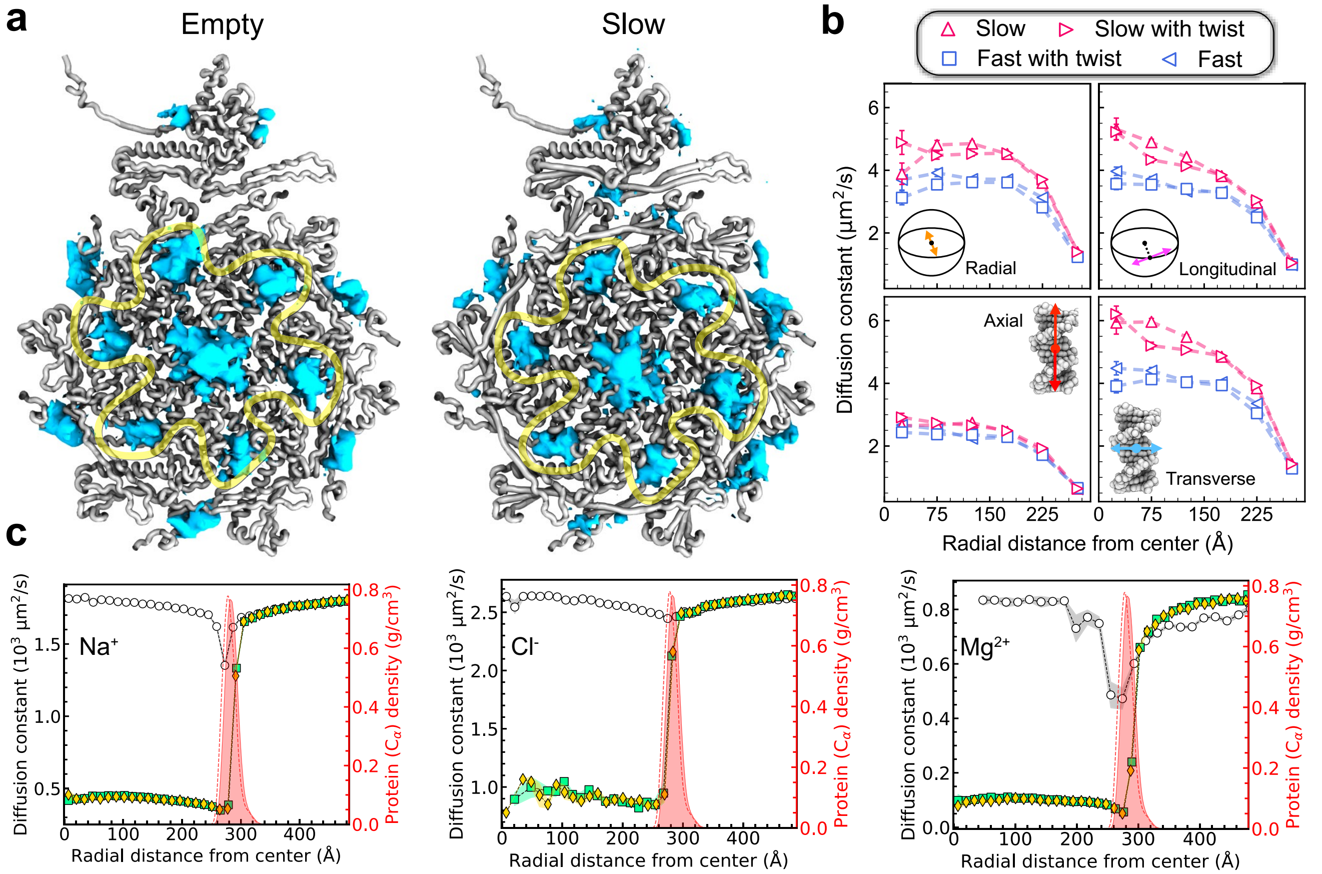
Extended Data Fig. 2 | Electrostatic properties of empty and packaged capsids. a, Ion exchange rate during equilibration of the empty (left), "slow" (center) and "slow with twist" (right) packaged capsids. Ion exchange data were collected every 10 ns. The lines shown were obtained from the data using a Savitzky-Golay filter with a 110 ns window. **b**, Radial profiles of the electrostatic potential averaged over 50 ns (empty; left) or 100 ns (slow and slow with twist; center and right) windows. The electrostatic potential was computed using VMD PMEpot plugin[91]. **c**, Image depicting the final result of an algorithm used to select the interior and exterior of the capsid, see Methods for details. **d**, Total charge inside the protein capsid in the units of proton charge. Traces are shown with (solid lines) and without (dotted lines) inclusion of the protein capsid in the analysis. Water molecules were neglected in the analysis. **e**, Charge inside a spherical volume centered at the center of the capsid as a function of the sphere's radius. Data were averaged over the last 100 ns of each simulation. For reference, the density of C$_\alpha$ atoms is shown (blue histogram; right axis). **f**, Electrostatic dipole moment of an asymmetric subunit projected along the radial axis. At each frame, the moment was averaged over all sixty copies of the asymmetric subunit. For reference, the blue line shows the average dipole moment of the residues resolved in the X-ray structure. **g**, Theoretical model used to estimate the electrostatic potential across the capsid (see Methods). **h**, Average potential difference between solvent occupied regions inside and outside the capsid as computed using PMEpot (circles) and from a fit using the theoretical model (dashed lines). In contrast to analysis shown in Fig. 3i, j, the PMEpot averaging was restricted to the solvent region within the capsid. That was done by blurring the mass density of DNA nucleobases via convolution with a 1 nm wide gaussian kernel and then selecting regions where the blurred density was less than 0.05 Da Å^{-3}. The fit was performed by minimizing the MSD between predicted and measured data points, yielding the following values for the parameters of the model: $q_{\rm extra,empty}$ = 248.4 e, $q_{\rm extra,slow}$ = 404.9 e, and $\varepsilon_{\rm pro}$ = 6.53.



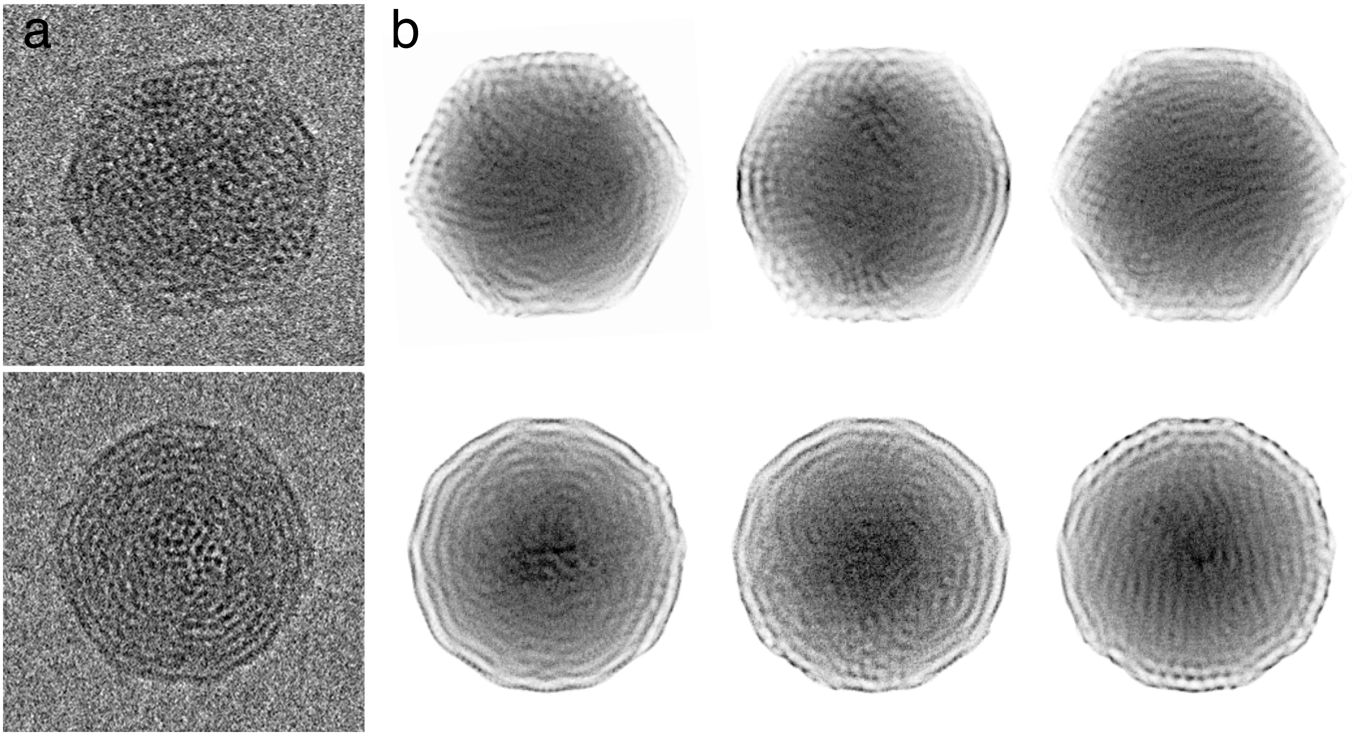
Extended Data Fig. 3 | Properties of packaged genome configurations. **a**, Sixteen genome configurations obtained by independent coarse-grained packaging simulations performed under the same 55 pN packaging force in the absence of twist (left) and with a 14°-per-10-bp twist imposed by the packaging protocol (right). Each genome configuration has a unique interface between the early- (blue) and late-packaged (red) DNA domains. **b**, Statistical properties of the packaged DNA configurations. Each plot depicts the distributions of a locally-defined metric for beads assigned in five different radial ranges, each with a 5 nm width, except the innermost group, which includes all beads within 7.5 nm of the capsid center. The distributions are semi-transparent and are provided for each packaged capsid. The mean of each distribution is shown as a solid horizontal line segment. The first three columns of plots characterize the angle between the local tangent at each bead and the spherical basis vectors at the bead's location. The last three columns depict local curvature, toroidal order[27,51] and nematic order. The fast packaged capsids exhibit significant differences according to several of the metrics. For example, the fast packaged DNA located within 7.5 nm of the capsid center points away from the portal by 5° on average, whereas the slow packaged DNA points towards and away from the portal with roughly equal likelihood. Compared to the slow packaged genomes, the fast packaged DNA has a slightly higher average curvature, lower nematic order, and, especially when twist was not imposed, greater toroidal order.



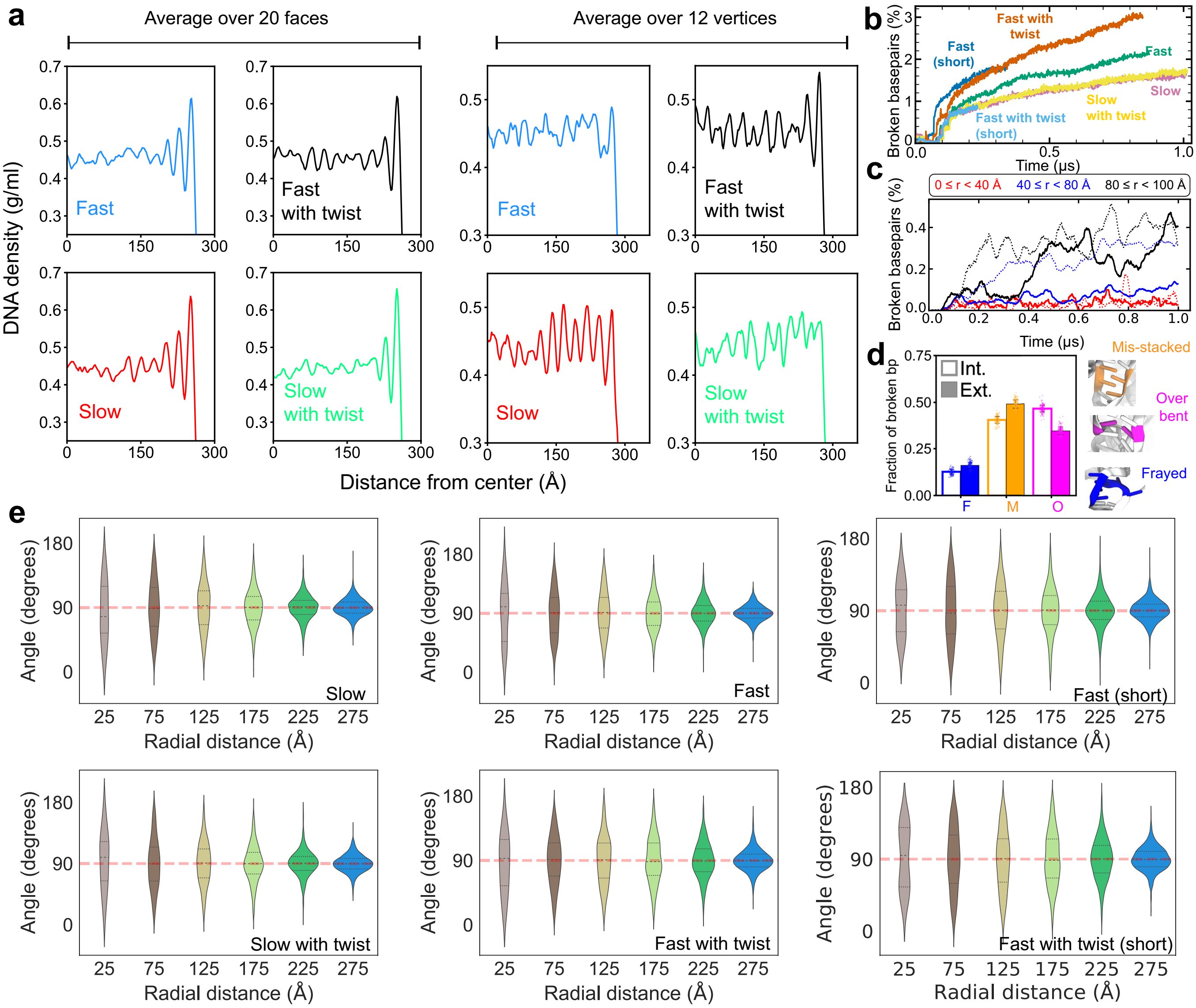
Extended Data Fig. 4 | Structural features of packaged capsids. **a**, Trajectory-average distance from the center of the capsid to the CoM of each capsid protein versus the corresponding crystal structure value. For each protein location, the symbol shows the value averaged over the 60 copies of the protein, whereas the histogram (top axis) shows the distribution among the protein copies. Data are shown from the last 50 ns of each trajectory, sampling coordinates every 0.192 ns. **b**, Average per-residue RMSD of the "slow" and "slow with twist" capsids from the crystallographic coordinates as a function of the residue number. RMSD of modeled residues is not shown. The averaging was done over the last 105 ns of each trajectory, sampling coordinates every 0.192 ns. The proteins at locations 1 to 5 exhibit consistent patterns (left), and the average of these data is depicted on the right (black). **c**, Average per-residue displacement of the packaged capsid with respect to the empty capsid. Deviation of modeled residues is not shown. The averaging was done over the last 40 ns of each trajectory, sampling coordinates every 48 ps. The mean and standard deviation over protein locations 1 to 5 are shown in black and grey, respectively.



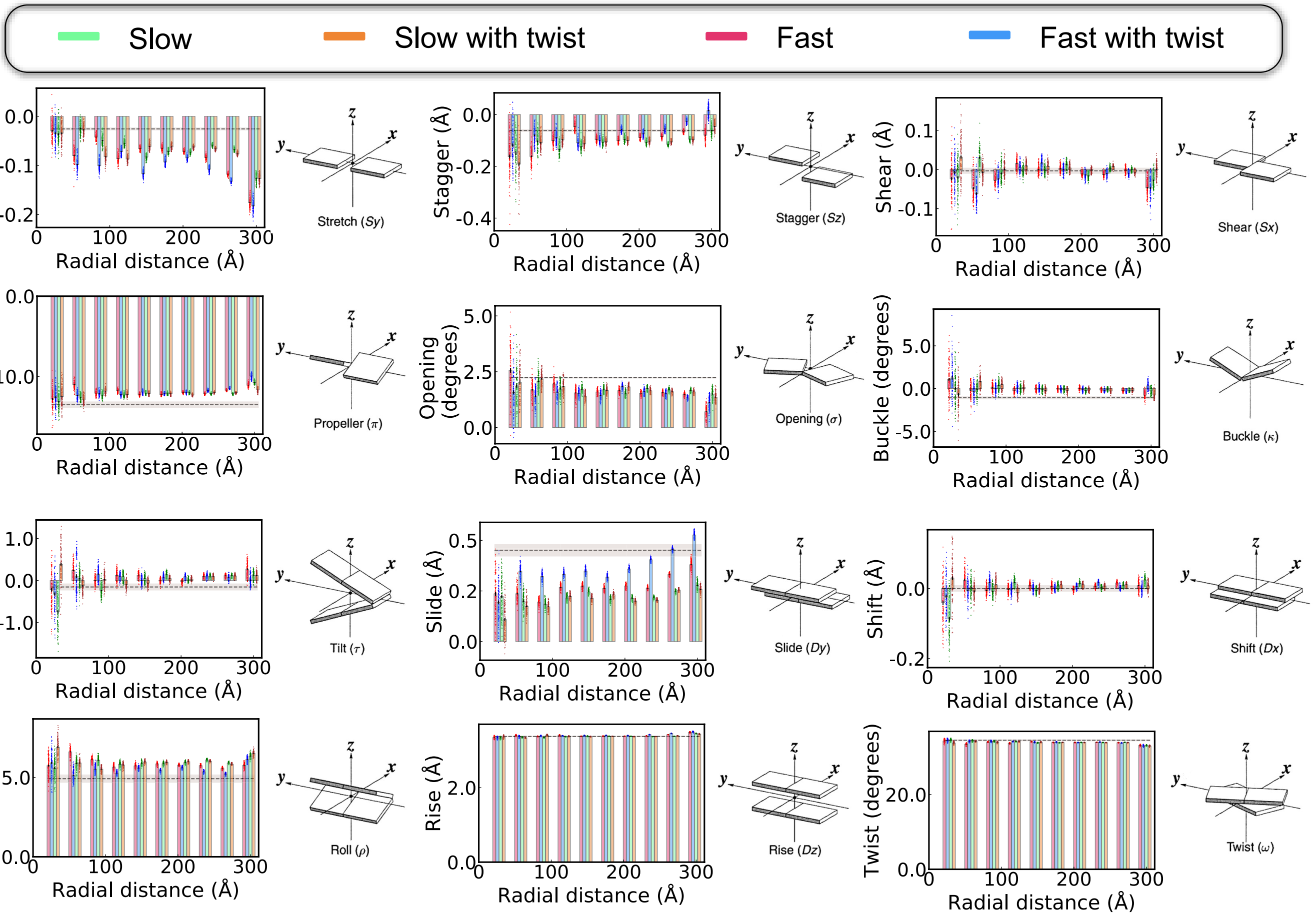
Extended Data Fig. 5 | Supplementary analysis of the all-atom MD trajectories. a, Location of water passages (blue molecular surface) within the capsid visualized as an isosurface (5.26 molecules nm^{-3}) of water density, averaged over the 60 icosahedron subunits, for the empty (left) and packaged (right) capsids. The analysis was performed using the last 2 ns of the each MD trajectory. The data in the left image are the same as in Fig. 1i. The ratio of the volume occupied by water (slow to empty) is 0.68. Less prominent differences in the water passages are observed inside the region confined by the yellow semi-transparent curve. Prominent differences are observed outside that region, where adjacent asymmetric subunits meet. Thus, expansion of the capsid caused by the packaged DNA leads to reduction of the gaps between the adjacent protein subunits. **b**, Diffusivity of DNA helices plotted as a function of their radial distance from the capsid's center. Error bars reflect standard deviation across all points within each radial bin. Data shown in the upper panels were obtained in the reference frame of the capsid whereas those in the lower panels were obtained in the local helical reference frame. The insets schematically illustrate the motion characterized by the corresponding diffusion coefficients. **c**, Radial distribution of ion diffusivity (left axis). Symbols correspond to the simulation models: empty circles for "empty", yellow diamonds for "slow with twist" and green squares for "slow", respectively. The radially averaged protein density (right axis) is shown for the packaged (filled distribution) and empty (dashed line) simulation trajectory.



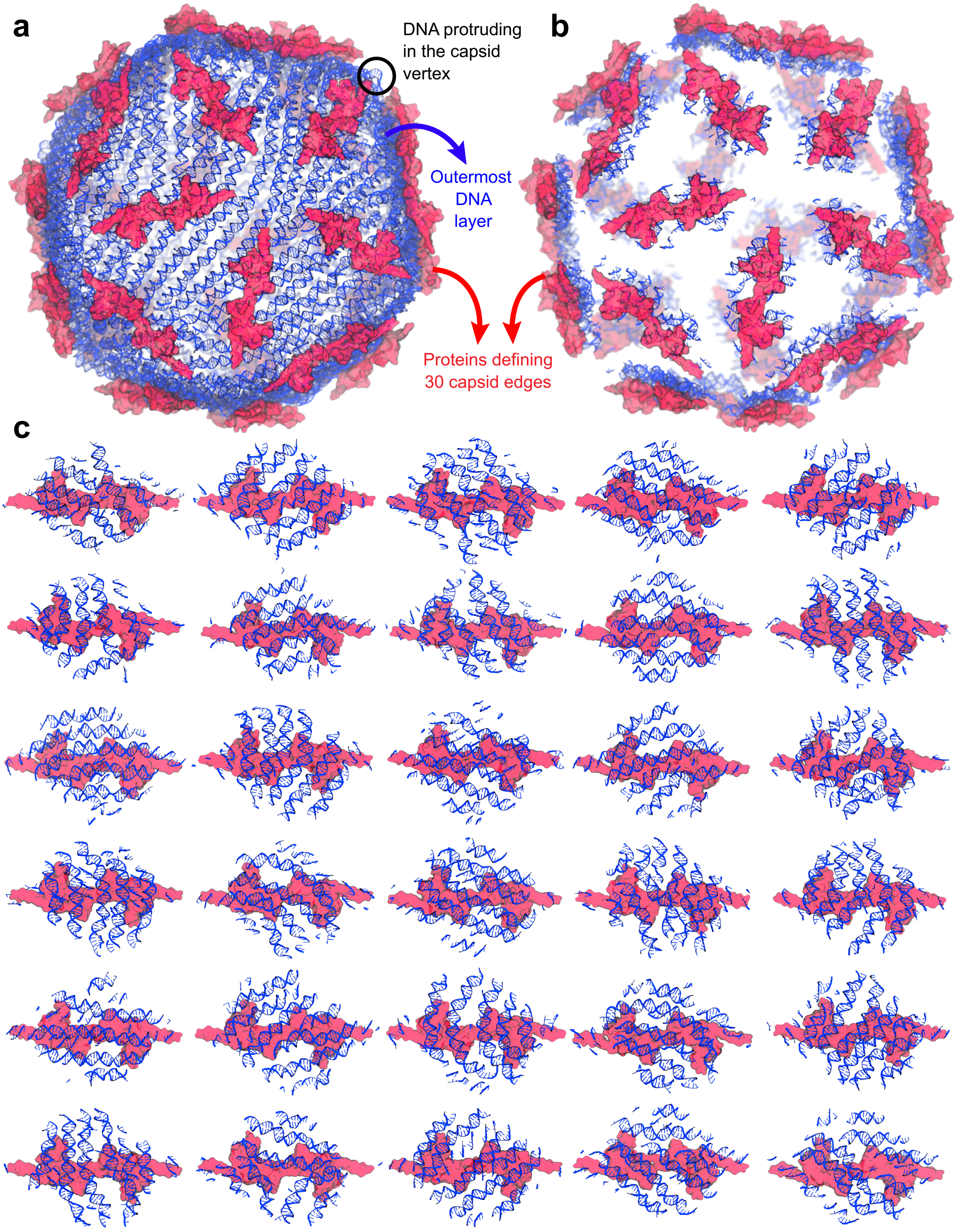
Extended Data Fig. 6 | TEM-like analysis of packaged HK97 genomes. **a**, TEM images of HK97 viral particles. **b**, TEM-like images of computationally packaged HK97 genomes. To make the TEM-like images, DNA mass density obtained from the coarse-grained packaging simulations was projected along several axes for several packaged genome configurations. The capsid density was not included in the analysis. Image in **a** reproduced from ref. 17.



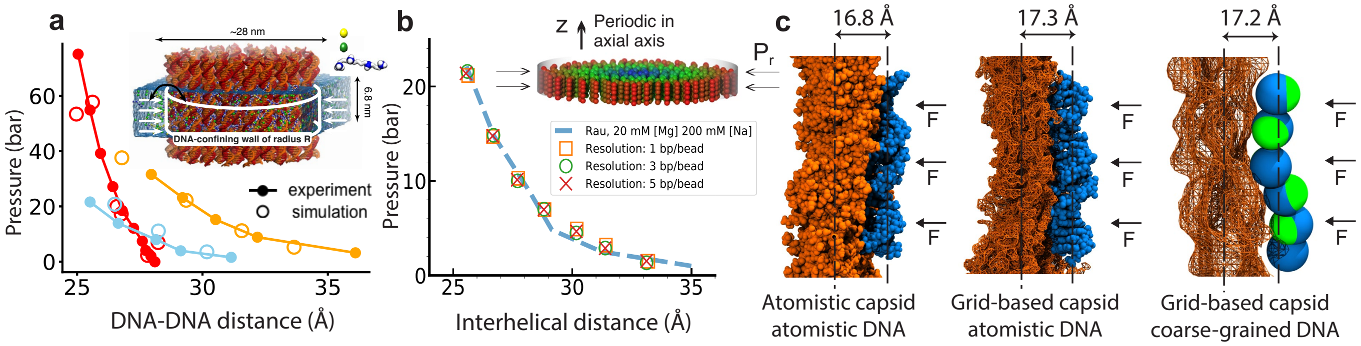
Extended Data Fig. 7 | Properties of packaged genomes according to all-atom MD simulations. **a**, Simulated profiles of DNA density for the four microsecond-long all-atom MD trajectories. The simulated profiles were computed by averaging 4 × 4 nm^2 centered sections connecting the ten pairs of opposite faces (left) or the six pairs of opposite vertices (right) and by averaging over the last 500 ns of the respective trajectories. The conformations resulting from the slow packaging simulations show greater DNA ordering (more visible layers) compared to the conformations resulting from fast packaging simulations. Lower but persistent order is observed along the vertices' symmetry axes. The configurations sampled by the simulations of the "slow" packaged particle show a higher ordering of the DNA near vertices compared to other simulations. The variations observed in individual structures may arise from the inherent stochasticity associated with the process of packaging. **b**, Fraction of base pairs broken in the DNA genome during the equilibration simulations of the six packaged particles. A base pair is considered intact if the H1 or N1 atom of a purine is within 2.5 Å of the N3 or H3 atoms of a pyrimidine, and the angle formed by the N1-H1-N3 or N1-H3-N3 atoms is greater than 115 degrees. **c**, Fraction of base pairs broken within three internal radial bins, analyzed every 0.96 ns of the "slow" (solid lines) and "slow with twist" (dotted lines) trajectories. **d**, Fraction of broken basepairs characterized according to their conformation (frayed, mis-stacked and over bent), analyzed over the last 50 ns of the "slow" trajectory. Exterior (Ext.) refers to DNA base pairs that have at least one non-hydrogen atom within 20 Å of the protein non-hydrogen atom, and interior (Int.) refers to all other base pairs. Error bars reflect standard deviation across all points within each radial bin. **e**, Violin plots of the angle between the local axis of a 10 bp DNA fragment and a radial vector as a function of the radial distance to the capsid center. The CoM of each 10 bp fragment was recorded for the last 288 ns of the equilibration trajectory, sampled every 0.48 ns. As expected, helices tend to align transverse to the radial vector, as one moves from the capsid center to the outermost layer (blue). A relatively higher bimodality is observed when packaging is performed at higher force in the bins next to the outermost layer, i.e. the two bins shown in green.



Extended Data Fig. 8 | Base pair-level characterization of the all-atom genome structures. The analysis was performed by first writing down separate coordinate files for every 150 bp of the genome for the last 5 ns of each all-atom trajectory every 48 ps. Each individual coordinate file was then analyzed using the Curves+ package[86]. The base-pair level properties were then averaged according to the base pairs' radial distance from the capsid center and normalized with respect to the number of base pairs within each radial bin and error bars depict s.e.m. over $n = 104$ consecutive segments of the trajectory. Colors indicate the different packaged models. Dashed line depicts mean deviation for two DNA duplexes having a random sequence of 28 bp in a 0.1 mol kg^{-1} KCl solution, each simulated for 90 ns in the NPT ensemble. For both systems, base-pair parameters were averaged separately for last two 40 ns intervals, giving four independent samples for computing the standard error. Schematic images adapted from ref. 59, Springer Nature.



Extended Data Fig. 9 | Topological defects in the structure of DNA genome near the edges of the capsid. a, Arrangement of the DNA molecules (blue) in the outermost layer of the genome at the end of the all-atom MD equilibration of the packaged capsid (slow trajectory). Proteins forming the capsid edges are shown in pink; the rest of the assembly is not shown for clarity. **b**, Same as in the previous panel, showing only the DNA helices located within 20 Å of the capsid edges. **c**, For each capsid edge (pink) the DNA helices from the previous panel are separately shown, viewed from the inside to outside.



Extended Data Fig. 10 | Multi-resolution model of DNA–DNA and DNA–protein interactions. a, Explicit solvent all-atom MD simulations of internal pressure in a DNA array[50]. Color indicates bulk electrolyte molarity: 20 mM Mg^{2+}/200 mM Na^+ (cyan), 250 mM Na^+ (orange), and 2 mM Sm^{4+}/200 mM Na^+ (red). Using CUFIX corrections to non-bonded interactions[63] was essential to achieve quantitative agreement with experiment. **b**, CG simulations of DNA array pressure at multiple resolutions. The internal pressure matches experimental values regardless of the resolution of the model. The CG simulations were performed using the mrDNA model[77]. **c**, Calibration of DNA–protein interactions. In each simulation, a DNA molecule was pushed against a flat cross-section of the viral capsid by an external force corresponding to a 20 bar pressure. The grid-based representation of the protein capsid was tuned to match the average DNA–protein distance seen in the all-atom simulation. Image in **a** adapted from ref. 50, Oxford Univ. Press.

nature portfolio

Corresponding author(s): Aleksei Aksimentiev

Last updated by author(s): Dec 18, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Multi-resolution (MR) packaging simulations were performed using MrDNA (version 1.0a), available at: https://gitlab.engr.illinois.edu/tbgl/tools/mrdna MrDNA is based on the ARBD (Atomic Resolution Brownian Dynamics) engine. The packaging simulations were performed using a developmental version based on the Oct 2019 beta release version, and we have supplied the source code in the IDB repository: https://doi.org/10.13012/B2IDB-4930709_V1 Parameters for the atoms affected by cross-linking were obtained from the CGenFF server (version 2.2.0). All-atom simulation trajectories were obtained using a memory optimized version of NAMD2.13b.
Data analysis	All simulation systems were constructed and visualized using VMD 1.9.4a43. Custom code, either using Tcl 8.0 or Python 3.7.3, was used to analyze the MD trajectories. For reading the trajectory (DCD) files, either VMD (for tcl) or MDAnalysis 1.1.1 (for Python) was used. Volumetric densities were generated by either VMD or the scikit-image library in Python. All visualizations were done with VMD. Small-angle X-ray scattering (SAXS) profiles of simulated structures were computed using Crysol 2.8.3.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Simulation trajectories corresponding to the main text figures will be available via a DOI, prior to publication. The datasets supporting the plots generated during the current study are attached. Any other data and simulation trajectories are available upon request.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The number of MD trajectories was determined by the maximum number of runs that could be performed using a state-of-the-art suprcomputer system
Data exclusions	No data, obtained from the simulation trajectories, were excluded in the analyses
Replication	Multiple replica simulations (from 2 to 8) were performed for each conditions. All replica simulations were successful
Randomization	Each replica simulation was initiated from a unique, randomly chosen microscopic configurations. Additional randomization was introduced through stochastic forces of the thermostat
Blinding	N/A. The very nature of computational experiments makes it impossible to implement traditional blinding. Replica simulations fulfilled that role as the outcome of the simulation was not known a priori

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*