Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

All-atom molecular dynamics of the cyanopodophage A4 portal-tail complex reveals dynamic asymmetry in tail fiber chains

Kirill Medvedev

kirill.medvedev@ucf.edu

University of Central Florida

Article

Keywords:

Posted Date: June 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9763141/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Cyanophages are key regulators of cyanobacterial populations in aquatic ecosystems, yet little is known about the conformational dynamics of their infection machinery. We performed a 250 ns all-atom molecular dynamics simulation of the cyanopodophage A4 portal-tail complex (PDB: 9K09), a 48-chain assembly of approximately 4 million atoms in explicit solvent. The simulation revealed two distinct groups: a rigid scaffold formed by the portal, adaptor, and nozzle, and significantly more flexible tail fibers. Within each tail fiber, the most mobile regions are in the C-terminal cell wall binding domain (CBD), which recognizes the *Anabaena* lipopolysaccharide receptor, while the disordered central stem is stabilized by inter-chain contacts. The three tail fiber chains in each trimer are sequence-identical, but their dynamics differ by position: the chain whose CBD points toward the capsid is more mobile than the other two. These results provide the first dynamic characterization of cyanopodophage A4.

1. Introduction

Cyanophages are crucial regulators of cyanobacterial populations in aquatic ecosystems and play central roles in global biogeochemical cycles through their effects on primary production and nutrient cycling[1]. Among the major morphotypes of tailed cyanophages, the podophages are characterized by short non-contractile tails that are too short to span the bacterial cell envelope[2]. Unlike myophages and siphophages that inject DNA through tail tubes or tape measure proteins, podophages eject internal core proteins (pre-ejectosome) from the capsid through the portal-tail complex, which then assemble into a transmembrane channel (ejectosome) required for genome delivery into the host cell[3]. While recent advances in cryo-EM have revealed the architecture of these molecular machines at near-atomic resolution, including the complete structure of cyanopodophage A4 with its pentameric pre-ejectosome[4], these static snapshots cannot capture the conformational dynamics essential for understanding the infection mechanism.

The tail fibers of tailed phages serve as the primary host recognition apparatus and must balance structural stability with conformational flexibility needed to scan host surfaces for specific receptors[5]. Engineering studies have demonstrated that tail fiber tip regions exhibit position-dependent mobility, with surface-exposed loops tolerating mutations while maintaining overall structural integrity[6]. Despite their functional importance, we still know little about how tail fiber assemblies move within intact phage structures: whether the chains fluctuate independently or together, and how each chain's position within that arrangement shapes its mobility.

Molecular dynamics (MD) simulations have become indispensable for bridging the gap between static structures and functional dynamics in biological systems. In recent years, computational virology has witnessed remarkable progress. All-atom (AA) MD simulations can now tackle systems of unprecedented size and complexity, though these studies remain computationally demanding tasks requiring high-performance computing resources. Advances in computing architectures and methodologies now make it possible to construct and simulate biological systems at unprecedented

scales, from complete viral capsids to respiratory aerosols[7]. Modern simulations now reach system sizes from millions to over a billion atoms, enabling all-atom modeling of complete virions and viral machines rather than their isolated parts[8]. Recent studies demonstrated the feasibility of simulating complete viral capsids, beginning with satellite tobacco mosaic virus[9] and expanding to larger assemblies including HIV-1[10], influenza[11], and SARS-CoV-2[12]. MD simulations uniquely capture dynamic domain movements that are essential for viral function yet invisible in static structures. Sztain *et al.*[12] employed over 130 μs of weighted ensemble simulations to reveal the complete atomic-resolution pathway of SARS-CoV-2 spike protein RBD opening – a large-scale domain movement (~ 20 Å displacement) from a glycan-shielded "down" to an exposed "up" conformation required for receptor binding and viral entry. Detecting this domain motion was an important step, since it identified the glycan as a functional regulator rather than passive shield, directly informing experimental mutagenesis that confirmed its gating role in receptor binding. Overall, domain movements in viral assemblies regulate essential functions including receptor binding, membrane fusion, and immune evasion. Their characterization through MD simulation provides mechanistic insights that complement static structures and inform therapeutic strategies[11–13].

Despite these advances, MD simulations of complete phage tail assemblies remain scarce, particularly for cyanophages. The cyanopodophage A4 portal-tail complex represents an ideal system for such investigations: 48 protein chains organized into portal, adaptor, nozzle, and tail fiber modules, totaling 21,978 residues that yield a system of approximately 4 million atoms when solvated in explicit water. Here we present a 250 ns all-atom MD simulation of the complete cyanopodophage A4 portal-tail assembly[4] to characterize the dynamics of its proteins and elucidate how trimeric architecture influences tail fiber mobility. Our analysis revealed that the portal, adaptor, and nozzle form a rigid assembly that contrasts with the significantly more flexible tail fibers. Within each tail fiber, the most mobile regions localize in the C-terminal cell wall binding domain that mediates host recognition, while a long, mostly disordered stem region remains constrained by inter-chain contacts. Despite being sequence-identical, the three chains of each trimer behave differently in our simulation: the chain facing the capsid is the most mobile, while the other two move similarly to one another.

2. Materials and Methods

System Preparation for Molecular Dynamics

We obtained the experimental structure of the cyanopodophage A4 portal-tail complex[4] from the Protein Data Bank (PDB: 9K09) and prepared it for simulation using GROMACS 2025.2[14]. The structure was parameterized using the AMBER99SB-ILDN force field[15], which has proven reliable for large protein assemblies like this one. We added missing hydrogen atoms at pH 7.0 and centered the protein in a triclinic box, leaving at least 2.4 nm of clearance between the protein surface and box edges to avoid artifacts from the periodic boundaries. After solvating the box with TIP3P water[16], we added sodium and chloride ions to both neutralize the system. This gave us a final system of roughly 4 million atoms:

333,594 protein atoms across 48 chains (21,978 residues total), approximately 3.6 million water atoms (around 1.2 million TIP3P molecules), and 7,524 ions.

Energy Minimization

Before dynamics, we minimized the potential energy using steepest descent for up to 50,000 steps until the maximum force fell below 1000 kJ mol^{-1} nm^{-1}. This relatively loose convergence criterion is appropriate for large systems where the primary goal is removing steric clashes introduced during solvation, with further structural relaxation occurring during the multi-stage equilibration protocol. Electrostatics were handled with Particle Mesh Ewald (PME)[17] using a 1.2 nm real-space cutoff and 0.16 nm grid spacing. Van der Waals interactions used a 1.2 nm cutoff. All hydrogen-containing bonds were constrained with LINCS.

Equilibration Protocol

Given the system's large size (~ 4 million atoms), we used a six-stage equilibration protocol spanning 2.5 ns. First, two NVT stages (300 ps each) brought the system to the target temperature of 310 K using the velocity-rescale thermostat[18]. We coupled protein and solvent separately with a 0.1 ps time constant and generated initial velocities from a Maxwell-Boltzmann distribution. During NVT, protein heavy atoms were position-restrained to prevent structural distortions during initial heating. Next, four NPT stages (500 ps each) equilibrated the system pressure. We started with the Berendsen barostat[19] in NPT1 for gentle initial relaxation, then switched to the more rigorous C-rescale barostat[20] for NPT2-4. All NPT stages maintained 1.0 bar pressure with isotropic coupling. Position restraints were weakened during NPT2-3, then completely removed for the final NPT4 stage. This last 500 ps of unrestrained equilibration ensured the system reached a stable state before production.

Production Run

We ran 250 ns of unrestrained production MD at 310 K and 1 bar and ultimately analyzed only 200 ns (from 50–250 ns) based on convergence of RMSD and radius of gyration. The simulation used a 2 fs timestep enabled by constraining all hydrogen bonds with LINCS. Temperature control used the velocity-rescale thermostat (τ = 0.1 ps) and pressure control used C-rescale (τ = 2.0 ps), both of which have been shown to produce correct thermodynamic ensembles. Electrostatics were calculated with Partial Mesh Ewald (PME) using the same settings as minimization. To manage storage for this large system, we saved coordinates every 100 ps.

Analysis

We analyzed trajectories using standard GROMACS tools supplemented with custom Python scripts. Structural stability was assessed via backbone RMSD (calculated for N, Cα, C, and O atoms after least-squares fitting) and radius of gyration. Per-residue flexibility was quantified using root mean square fluctuation (RMSF) analysis. RMSF values were computed for all Cα atoms across the 48-chain complex

(21,978 residues total) using GROMACS 2025.2 gmx rmsf module over the production simulation period (50–250 ns). Residue numbering was renumbered sequentially across all chains for visualization. To reduce noise and reveal domain-level flexibility patterns, RMSF values were smoothed using a uniform filter with a 100-residue sliding window. The smoothed value at each position i is calculated as the arithmetic mean of RMSF values within a window of k = n/2 = 50 residues on each side:

$$RMSF_{smooth}(i) = \frac{1}{n} \sum_{i+k-1}^{j=i-k} RMSF_{raw}(j)$$

where n = 100 is the window size and k = 50 is the half-window radius. This window size balances noise reduction. Both raw and smoothed data are presented.

For per-chain analyses, residues were grouped into their respective chains based on the chain assignments in the topology file. For each of the 48 chains (designated A-R, U-Z, a-r, u-z according to PDB 9K09 nomenclature), mean RMSF, median RMSF, standard deviation, and range (minimum and maximum values) were calculated from the per-residue RMSF data. Chains were classified by flexibility using the following thresholds: rigid (mean RMSF < 1.0 Å), moderately flexible (1.0–2.0 Å), and highly flexible (> 2.0 Å). The percentage of residues falling into each category was computed for each chain.

For each of the 18 tail fiber chains, we identified the tip residue as the Cα atom most distant from the portal center of mass in the equilibrated starting frame of the production trajectory. We then computed the Euclidean distance between this tip residue and the portal center of mass at each frame of the 50–250 ns production window, producing one tip-to-portal distance trajectory per chain. The standard deviation of each chain's distance trajectory was used as a per-chain mobility metric, capturing the amplitude of fluctuations around each chain's mean position. Chains were grouped by structural position within the trimer (proximal-to-capsid, interior, distal; 6 chains each). Mobility values were compared between groups using independent two-sample t-tests, with Bonferroni correction for three pairwise comparisons (adjusted α = 0.017).

3. Results and Discussion

3.1. Domain classification of cyanopodophage A4 portal-tail complex.

Classification of structural domains of viral proteins is a long-standing challenge, driven by the rapid pace of viral evolution and the resulting uniqueness of viral proteomes. As a result, many experimentally determined viral structures in the Protein Data Bank lack systematic classification of their constituent domains within evolutionary hierarchies such as ECOD, particularly for large multi-protein assemblies

that contain dozens of distinct chains. The cyanopodophage A4 portal-tail complex is no exception: while Hou *et al.*[4] identified several domains through structural comparisons to the T7 podophage, a comprehensive classification of all constituent domains has not been undertaken. Protein dynamics vary both between protein types in an assembly and between domains within a single chain. Domain classification enables analysis at the second, finer-grained level. Annotating domains lets us connect the position-dependent mobility we observe to the functional roles of each module. The cyanopodophage A4 portal-tail complex (PDB: 9K09) is a heteromeric 48-mer composed of four unique protein chains, each forming a different structural component: portal, adaptor, tail fiber, and nozzle[4]. We utilized AlphaFold3[21] to create models of the four proteins constituting the portal-tail complex and generate the Predicted Aligned Errors (PAE) associated with each model, which are required for further domain classification pipeline. We applied the Domain Parser for AlphaFold Models (DPAM) pipeline[22] to classify domains of the cyanopodophage A4 portal-tail complex within Evolutionary Classification of protein Domains (ECOD) hierarchy[23,24] (Fig. 1). The tail tubular protein A (UniProt: A0A059PY25), which forms the adaptor of the portal-tail complex, consists of two domains. The N-terminal domain (Fig. 1A, yellow) adopts an alpha arrays architecture and belongs to the Hypothetical protein YqbG-like ECOD homology group (H-group ID: 4275.1), while the C-terminal domain (Fig. 1A, green) adopts a beta barrels architecture and potentially belongs to the RIFT-related H-group (ID: 1.1). The tail tubular protein B (UniProt: A0A059PYE2) forms the nozzle of the complex and includes six distinct domains (Fig. 1B). Five out of six domains adopt beta sandwiches architecture: green (jelly-roll) belongs to the Galactose-binding domain-like H-group (ID: 10.32), blue (Baculovirus p35 protein-related H-group: 4081.1), wheat and yellow (Glycosyl hydrolase domain H-group: 12.1) and light orange (Putative glucosidase YicI H-group: 4178.1). The central domain (Fig. 1B, pink) is a 7-bladed beta propeller (H-group: 5.1). The portal protein (UniProt: A0A059PYA9) includes two domains (Fig. 1C): pink adopts alpha bundles architecture (Alpha-helical domain in upper collar protein H-group: 4038.1) and green adopts beta barrels architecture (Barrel domain in upper collar protein H-group: 4056.1). The disordered loop shown in Fig. 1C is not included into the experimental structure and was modelled by AlphaFold3. Finally, the tail fiber protein (UniProt: A0A059PY41) contains two beta-sandwich domains connected by a long, mostly disordered region (stem): an N-terminal domain (Fig. 1D, green; Agglutinin HPA-like H-group: 4182.1) and a C-terminal domain (Fig. 1D, yellow; TNF-like H-group: 10.3).

3.2. All-atom molecular dynamics simulation and structural stability of the cyanopodophage A4 portal-tail complex.

We performed a 250 ns all-atom MD simulation of the complete cyanopodophage A4 portal-tail complex (PDB: 9K09) in explicit solvent under physiological conditions, yielding a system of approximately 4 million atoms. Simulations were carried out in GROMACS 2025.2 with the AMBER force field, following standard energy minimization and a multi-stage equilibration protocol (see Materials and Methods). To assess the stability of the simulated complex, we monitored two global metrics throughout the

trajectory: backbone RMSD relative to the first frame of the production run (i.e., the structure obtained after energy minimization and multi-stage NVT/NPT equilibration of the experimental structure), and radius of gyration (Rg) as a measure of overall compactness (Fig. 2). Backbone RMSD increased rapidly during the first ~ 50 ns, reaching a plateau around 6.0-6.5 Å, and remained stable for the remainder of the trajectory (Fig. 2A). This initial deviation reflects the multi-million-atom assembly relaxing from its equilibrated starting configuration into a thermally equilibrated state at physiological temperature. The RMSD plateaus around 6.0 Å after 50 ns and fluctuates within ± 1 Å for the remainder of the trajectory (Fig. 2A). Together with the stable Rg (Fig. 2B), this confirms that the complex is suitable for production analysis over the 50–250 ns window. The radius of gyration shows a consistent pattern: starting at ~ 95.5 Å, Rg decreased gradually over the first ~ 75 ns and then stabilized in the 50–250 ns production window, fluctuating around ~ 93.9 Å (Fig. 2B). The small magnitude of the Rg change indicates that the overall architecture of the complex is stable throughout the simulation. Together, the RMSD plateau and the stable Rg confirm that the complex is structurally well-behaved throughout the production phase, providing a reliable basis for the flexibility analyses.

3.3. Global per-residue flexibility profile of the cyanopodophage A4 portal-tail complex.

With the complex shown to be structurally stable over the 50–250 ns production window, we next characterized its local flexibility by computing per-residue Cα RMSF values for all 21,978 residues across the 48 chains (Fig. 3A). The mean RMSF was 0.69 Å. Most residues fluctuated by 0.3–0.7 Å, while a smaller population of more mobile residues extended the distribution toward higher values (Fig. 3B). The majority of residues (81.2%) fall below the 1.0 Å rigid threshold (Fig. 3B), indicating that the complex behaves as a predominantly stable, well-packed assembly with only localized regions of elevated mobility.

The smoothed RMSF profile (Fig. 3A, dark red) reveals that flexibility is not uniformly distributed along the sequence. Residue numbering runs sequentially across all 48 chains (1 to 21,978 total), so peaks on the x-axis correspond to different chains in the assembly. Most of the complex fluctuates close to 0.5 Å, while five regions rise above the complex-wide mean: a smaller peak around residues 1-500, and four more pronounced hotspots around residues 7,500-9,000, 12,000–14,000, 16,500 – 18,000, and 20,500 – 22,000, reaching values of 1.5-2.0 Å. These regions indicate that increased mobility is concentrated in a specific subset of chains rather than being spread evenly across the assembly and motivate the chain- and residue-level analyses presented in the following sections.

3.4. Tail fiber chains are more flexible than the portal, adaptor, and nozzle proteins.

To determine which chains and protein types account for the flexibility hotspots observed in the global RMSF profile, we aggregated per-residue RMSF values by chain and ranked the 48 chains by mean

fluctuation (Fig. 4A). The chains separate cleanly into two groups. All 18 tail fiber chains (red) cluster at the top of the ranking, with mean RMSF values between 0.75 and 1.26 Å, every one of them above the complex-wide mean (0.69 Å, dashed line). The remaining 30 chains, which form the portal (blue), nozzle (green), and adaptor (orange), are grouped below the mean, with values between 0.42 and 0.64 Å. No tail fiber chain falls below the complex-wide mean, and no portal, nozzle, or adaptor chain rises above it.

Averaged by protein type (Fig. 4B), tail fiber proteins showed a mean RMSF of 1.06 Å, roughly two-fold higher than the portal (0.52 Å), nozzle (0.54 Å), and adaptor (0.48 Å). This pattern is consistent with the structural roles of the four protein groups of the complex: the portal, adaptor, and nozzle assemble into the short tail at the capsid vertex. The portal anchors the pre-ejectosome inside the capsid, while the nozzle retains the packaged DNA. Together, these three proteins form the channel through which ejection proteins and viral DNA exit during infection[4]. The tail fibers project outward from the nozzle and bear cell binding domains that specifically recognize and bind host surface receptors[4]. The flexibility hotspots identified in Fig. 3A therefore localize specifically to the tail fibers, motivating a closer look at their internal dynamics in the following sections. This dynamic partitioning matches the functional logic of the four proteins: components that gate genome handling must maintain defined geometry, while components that engage variable host surfaces benefit from conformational flexibility.

3.5. The most flexible regions of the tail fiber map onto the cell wall binding domain.

Zooming in from chains to residues, we examined the per-residue RMSF profile of chain Z, the most flexible chain in the assembly (mean RMSF 1.26 Å; Fig. 4A), to identify the region where increased mobility is observed. The RMSF profile of chain Z is uneven along the sequence (Fig. 5A): the N-terminal domain and central stem show comparable, moderate fluctuations, while the C-terminal domain stands out with substantially higher mobility. The central stem (residues 111–230), which is largely disordered in the experimental structure[4], fluctuates with comparable amplitude to the folded N-terminal domain (residues 1-110). This is notable because disordered regions in isolation usually exhibit substantially higher mobility than folded domains[25]. In our simulation, the stem's restrained dynamics likely reflect close contacts with the corresponding stems of the other two chains in the trimer. This quaternary structure stabilizes the stem's positional freedom without requiring it to adopt a secondary structure. The C-terminal beta-sandwich domain (residues 231–372), in contrast, contains several pronounced peaks reaching 2.5–3.8 Å, with the three most flexible loop regions 236–253, 290–296 and 336–342 (Fig. 5B, magenta).

This C-terminal domain (Fig. 5B) corresponds to the cell wall binding domain (CBD) identified by Hou *et al.*[4] based on structural homology with validated receptor binding domains of related phages, including the *Anabaena*-binding short tail fiber gp33 of cyanophage A-1(L), which binds the lipopolysaccharide (LPS) O-antigen on the outer membrane of *Anabaena* sp. PCC 7120[26]. The flexibility hotspots we observe

therefore localize precisely to the host-contact region of the tail fiber (Fig. 5B, magenta), suggesting that conformational mobility is concentrated where the phage engages the LPS receptor.

The increased flexibility seen along chain Z raises the question of whether all 18 tail fiber chains in the complex behave similarly, or whether their mobility depends on each chain's structural position within the assembled tail fiber. Inspection of the structure shows that the three CBDs of each trimer occupy structurally distinct positions relative to both the portal and the capsid (Fig. 5C). One CBD sits on top of the other two and points back toward the capsid. We refer to this as the proximal-to-capsid chain. A second CBD lies closest to the portal axis, buried within the trimer; we refer to this as the interior chain. The third CBD extends furthest from the portal, oriented outward. We refer to this as the distal chain. Across the six trimers of the assembly, each position is occupied by six chains, giving 18 chains in three groups of six. We tested whether this geometric arrangement affect the position-dependent dynamics, presented in the following section.

3.6. Tail fiber tips show position-dependent mobility within the assembled trimer.

To quantify the mobility of each tail fiber chain, we tracked the distance from the C-terminal tip of each of the 18 tail fiber chains to the center of mass of the portal over the 50–250 ns production window (Fig. 6A). The standard deviation of each chain's distance trajectory served as the per-chain mobility metric. Because the absolute tip-to-portal distance depends on each chain's position in the assembly, fluctuation amplitude around the chain mean is more meaningful as a mobility metric than absolute distance. Using the standard deviation of the tip-to-portal distance as a per-chain mobility metric (Fig. 6B), the 18 chains divide into three groups by position. Proximal-to-capsid chains (y, q, Y, Q, b, B) exhibit the largest fluctuations, with values ranging from 3.1 to 5.1 Å. Interior chains (Z, r, R, c, C, z) and distal chains (X, p, P, a, A, x) are less mobile, with fluctuations between 2.2 and 3.6 Å. Grouped by structural position (Fig. 6C), proximal-to-capsid chains have a mean tip mobility of 3.99 ± 0.66 Å, compared with 2.93 ± 0.24 Å for interior chains and 2.83 ± 0.48 Å for distal chains. Independent t-tests confirm that proximal-to-capsid chains are significantly more mobile than both interior chains ($p = 0.007$) and distal chains ($p = 0.010$), while interior and distal chains are statistically indistinguishable ($p = 0.69$). Both significant differences survive Bonferroni correction for three pairwise comparisons (adjusted $\alpha = 0.017$).

What stands out is that the three chains of each trimer are chemically identical, yet they do not behave the same way. The proximal-to-capsid chain shows elevated mobility, while the interior and distal chains move at comparable, lower amplitudes. The asymmetry therefore arises from geometric position within the assembled trimer rather than from sequence or chemistry. It is important to note that the capsid is not included in our simulated system (PDB: 9K09 is the portal-tail complex alone). In the intact virion, the proximal-to-capsid CBD lies close to the lower surface of the viral capsid[4], whereas the interior and distal CBDs point away from it. The absence of the capsid in our simulation could affect the proximal-to-capsid CBD, whose elevated mobility may partly reflect the absence of this neighboring constraint.

Conclusions

In this study, we conducted a 250 ns all-atom molecular dynamics simulation of the cyanopodophage A4 portal-tail complex (PDB: 9K09), a 48-chain assembly comprising approximately 4 million atoms in explicit solvent. To our knowledge, this is the first all-atom MD simulation of a cyanopodophage portal-tail complex, and the first simulation of any component of cyanopodophage A4.

The simulation revealed several layers of organization in the dynamics of the complex. Overall, the assembly remained structurally stable throughout the production phase, with backbone RMSD plateauing at ~ 6 Å and the radius of gyration converging to a narrow range. At the chain level, the 18 tail fiber chains were more flexible than the 30 chains of the portal, adaptor, and nozzle, with mean RMSF values approximately two-fold higher. This reflects the two different roles of these proteins: the portal components hold the assembly together, while the tail fibers extend outward to bind the host[4,26]. At the residue level, the most flexible regions of the tail fiber chain localize within its C-terminal beta-sandwich domain, which corresponds to the cell wall binding domain (CBD) responsible for host recognition. We also observed that the disordered central stem of the tail fiber chain moves no more than the folded N-terminal domain in our simulation, despite its disorder in the experimental structure. This phenomenon can be likely explained by inter-chain contacts within the trimer and illustrates the value of simulating the assembled complex rather than isolated chains. Within each tail fiber trimer, the three chains are sequence-identical, yet the mobility of the C-terminal CBD differs significantly depending on each chain's geometric position within the trimer. The proximal-to-capsid chain, whose CBD points toward the capsid in the intact virion, is more mobile than the interior and distal chains, which behave similarly to each other.

Beyond these specific findings, our results demonstrate the value of all-atom MD on the assembled cyanopodophage portal-tail complex. Several of the observations reported here – the rigidity of the disordered stem, the localization of flexibility within the CBD, and the position-dependent dynamics within the trimer – would not be accessible from single-chain models or from the static experimental structure alone.

Declarations

Data availability

All molecular dynamics simulation files supporting this study are available at Zenodo (https://doi.org/10.5281/zenodo.20275757).

Competing interests

The authors declare that there are no competing interests associated with the manuscript.

Acknowledgements

The authors acknowledge the Advanced Research Computing Center (ARCC) at The University of Central Florida (https://arcc.ist.ucf.edu/) for providing computational resources that have contributed to the research results reported within this paper.

Funding

The study is supported by The University of Central Florida College of Engineering and Computer Science (to K.E.M.).

References

1. Goldin, S., Hulata, Y., Baran, N. & Lindell, D. Quantification of T4-Like and T7-Like Cyanophages Using the Polony Method Show They Are Significant Members of the Virioplankton in the North Pacific Subtropical Gyre. *Front Microbiol* **11**, 1210 (2020). https://doi.org/10.3389/fmicb.2020.01210
2. Casjens, S. R. & Molineux, I. J. Short noncontractile tail machines: adsorption and DNA delivery by podoviruses. *Adv Exp Med Biol* **726**, 143-179 (2012). https://doi.org/10.1007/978-1-4614-0980-9_7
3. Swanson, N. A. *et al.* Cryo-EM structure of the periplasmic tunnel of T7 DNA-ejectosome at 2.7 A resolution. *Mol Cell* **81**, 3145-3159 e3147 (2021). https://doi.org/10.1016/j.molcel.2021.06.001
4. Hou, P. *et al.* Cryo-EM structure of cyanopodophage A4 reveals a pentameric pre-ejectosome in the double-stabilized capsid. *Proc Natl Acad Sci U S A* **122**, e2423403122 (2025). https://doi.org/10.1073/pnas.2423403122
5. Taslem Mourosi, J. *et al.* Understanding Bacteriophage Tail Fiber Interaction with Host Surface Receptor: The Key "Blueprint" for Reprogramming Phage Host Range. *Int J Mol Sci* **23** (2022). https://doi.org/10.3390/ijms232012146
6. Yehl, K. *et al.* Engineering Phage Host-Range and Suppressing Bacterial Resistance through Phage Tail Fiber Mutagenesis. *Cell* **179**, 459-469 e459 (2019). https://doi.org/10.1016/j.cell.2019.09.015
7. Dommer, A. C., Wauer, N. A., Marrink, S. J. & Amaro, R. E. All-atom virus simulations to tackle airborne disease. *Curr Opin Struct Biol* **92**, 103048 (2025). https://doi.org/10.1016/j.sbi.2025.103048
8. Lynch, D. L., Pavlova, A., Fan, Z. & Gumbart, J. C. Understanding Virus Structure and Dynamics through Molecular Simulations. *J Chem Theory Comput* **19**, 3025-3036 (2023). https://doi.org/10.1021/acs.jctc.3c00116
9. Freddolino, 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). https://doi.org/10.1016/j.str.2005.11.014
10. Yu, A. *et al.* Strain and rupture of HIV-1 capsids during uncoating. *Proc Natl Acad Sci U S A* **119**, e2117781119 (2022). https://doi.org/10.1073/pnas.2117781119
11. Casalino, L. *et al.* Breathing and Tilting: Mesoscale Simulations Illuminate Influenza Glycoprotein Vulnerabilities. *ACS Cent Sci* **8**, 1646-1663 (2022). https://doi.org/10.1021/acscentsci.2c00981

12. Sztain, T. *et al.* A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nat Chem* **13**, 963-968 (2021). https://doi.org/10.1038/s41557-021-00758-3
13. Oliveira, A. S. F. *et al.* Allosteric modulation by the fatty acid site in the glycosylated SARS-CoV-2 spike. *Elife* **13** (2025). https://doi.org/10.7554/eLife.97313
14. Pall, S. *et al.* Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *J Chem Phys* **153**, 134110 (2020). https://doi.org/10.1063/5.0018516
15. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950-1958 (2010). https://doi.org/10.1002/prot.22711
16. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* **79**, 926-935 (1983). https://doi.org/10.1063/1.445869
17. Essmann, U. *et al.* A Smooth Particle Mesh Ewald Method. *Journal of Chemical Physics* **103**, 8577-8593 (1995). https://doi.org/10.1063/1.470117
18. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J Chem Phys* **126**, 014101 (2007). https://doi.org/10.1063/1.2408420
19. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A. & Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *Journal of Chemical Physics* **81**, 3684-3690 (1984). https://doi.org/Doi 10.1063/1.448118
20. Bernetti, M. & Bussi, G. Pressure control using stochastic cell rescaling. *J Chem Phys* **153**, 114107 (2020). https://doi.org/10.1063/5.0020514
21. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493-500 (2024). https://doi.org/10.1038/s41586-024-07487-w
22. Zhang, J., Schaeffer, R. D., Durham, J., Cong, Q. & Grishin, N. V. DPAM: A domain parser for AlphaFold models. *Protein Sci* **32**, e4548 (2023). https://doi.org/10.1002/pro.4548
23. Schaeffer, R. D. *et al.* ECOD: integrating classifications of protein domains from experimental and predicted structures. *Nucleic Acids Res* **53**, D411-D418 (2025). https://doi.org/10.1093/nar/gkae1029
24. Schaeffer, R. D. *et al.* ECOD domain classification of 48 whole proteomes from AlphaFold Structure Database using DPAM2. *PLoS Comput Biol* **20**, e1011586 (2024). https://doi.org/10.1371/journal.pcbi.1011586
25. Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* **16**, 18-29 (2015). https://doi.org/10.1038/nrm3920
26. Xiong, Z., Wang, Y., Dong, Y., Zhang, Q. & Xu, X. Cyanophage A-1(L) Adsorbs to Lipopolysaccharides of Anabaena sp. Strain PCC 7120 via the Tail Protein Lipopolysaccharide-Interacting Protein (ORF36). *J Bacteriol* **201** (2019). https://doi.org/10.1128/JB.00516-18

Figures

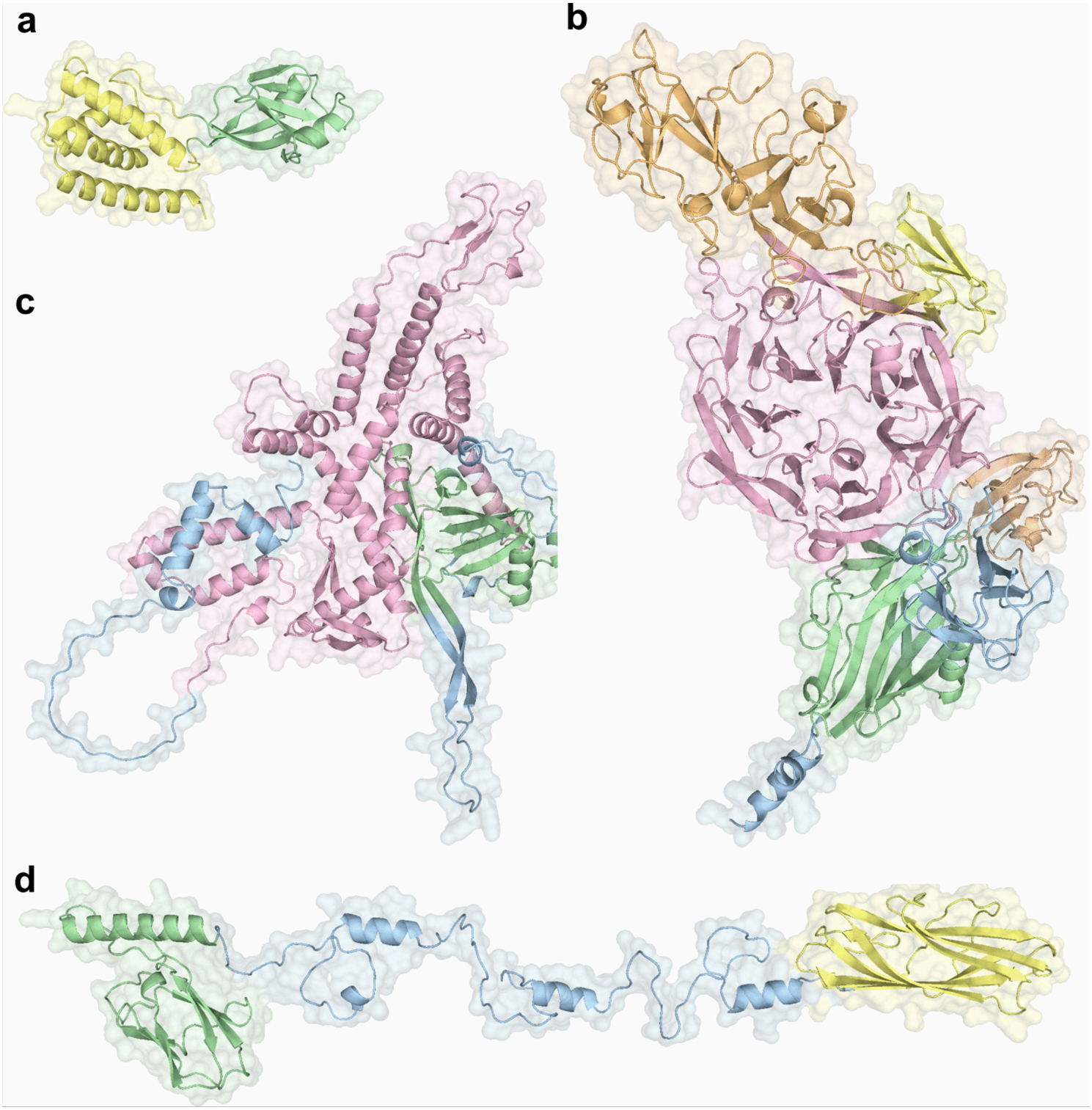


Figure 1

ECOD-based domain classification of the cyanopodophage A4 portal-tail complex. a Tail tubular protein A (UniProt: A0A059PY25). **b** Tail tubular protein B (UniProt: A0A059PYE2). **c** Portal protein (UniProt: A0A059PYA9). **d**Tail fiber protein (UniProt: A0A059PY41).

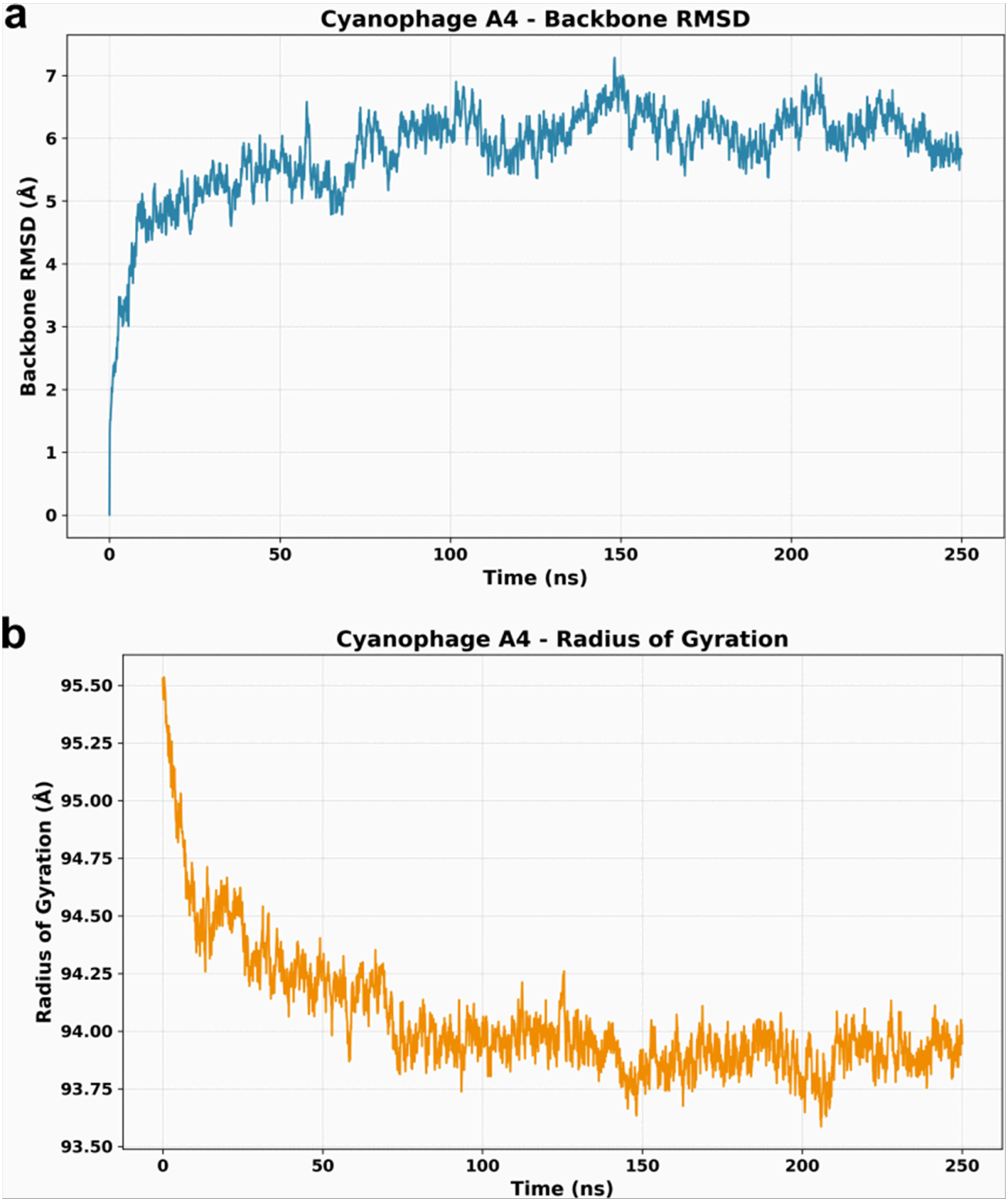


Figure 2

Structural stability metrics for the cyanopodophage A4 portal-tail complex during 250 ns MD simulation. a Backbone RMSD (N, Cα, C, O atoms) of the 48-chain complex relative to the first frame of the production trajectory, calculated after least-squares fitting on backbone atoms. **b** Radius of gyration (Rg) of the complex over the trajectory.

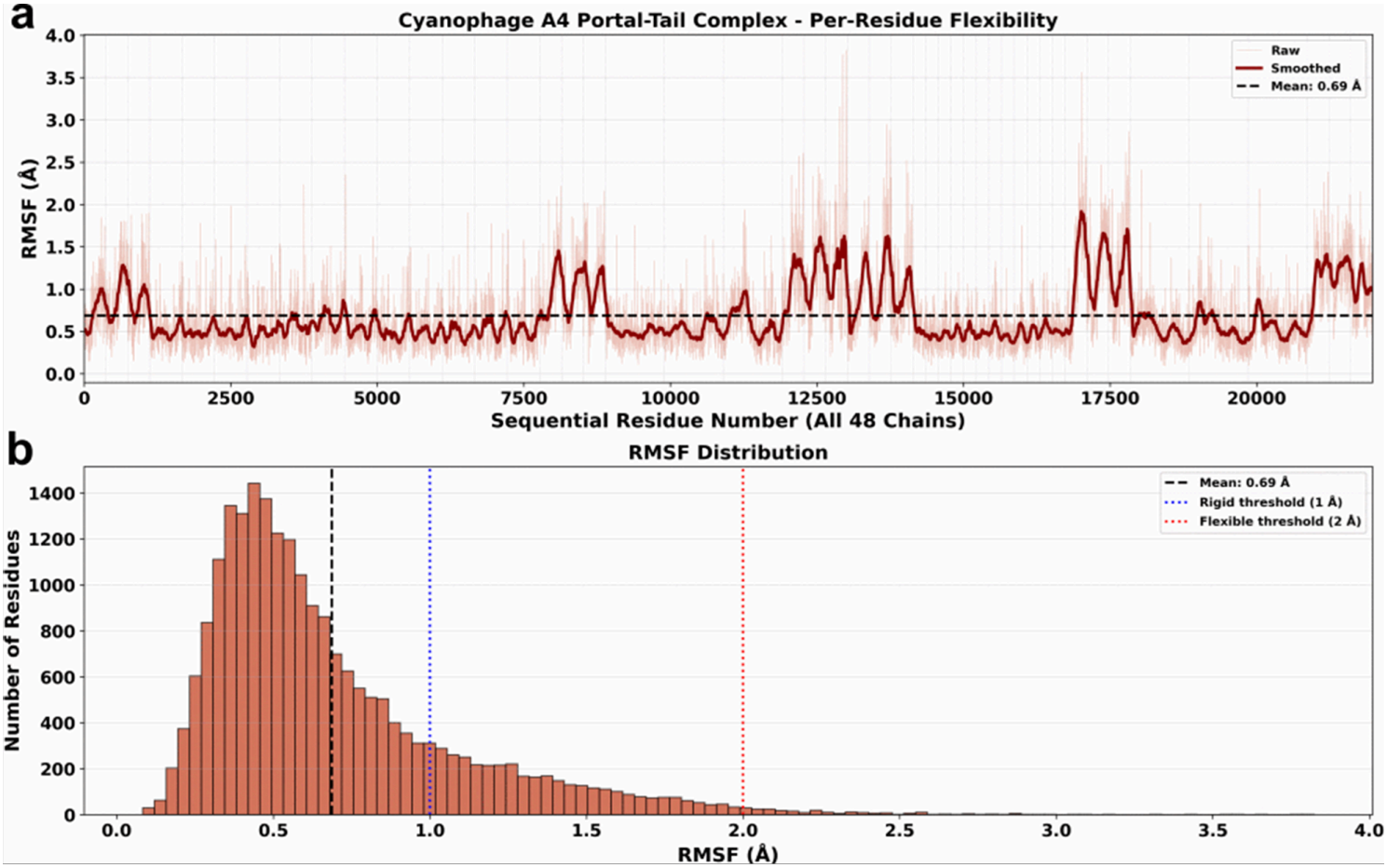


Figure 3

Global per-residue flexibility profile of the cyanopodophage A4 portal-tail complex. a Per-residue Cα RMSF for all 21,978 residues across the 48 chains, computed over the 50–250 ns production window. Residue numbering runs sequentially across all 48 chains (1 to 21,978 total). Light red: raw RMSF values; dark red: smoothed trace (100-residue moving-average filter). Gray vertical lines mark chain boundaries. Dashed line: complex-wide mean RMSF (0.69 Å). **b** Distribution of per-residue RMSF values across all 21,978 residues. Dashed line: mean RMSF (0.69 Å); blue dotted line: rigid threshold (1.0 Å); red dotted line: flexible threshold (2.0 Å).

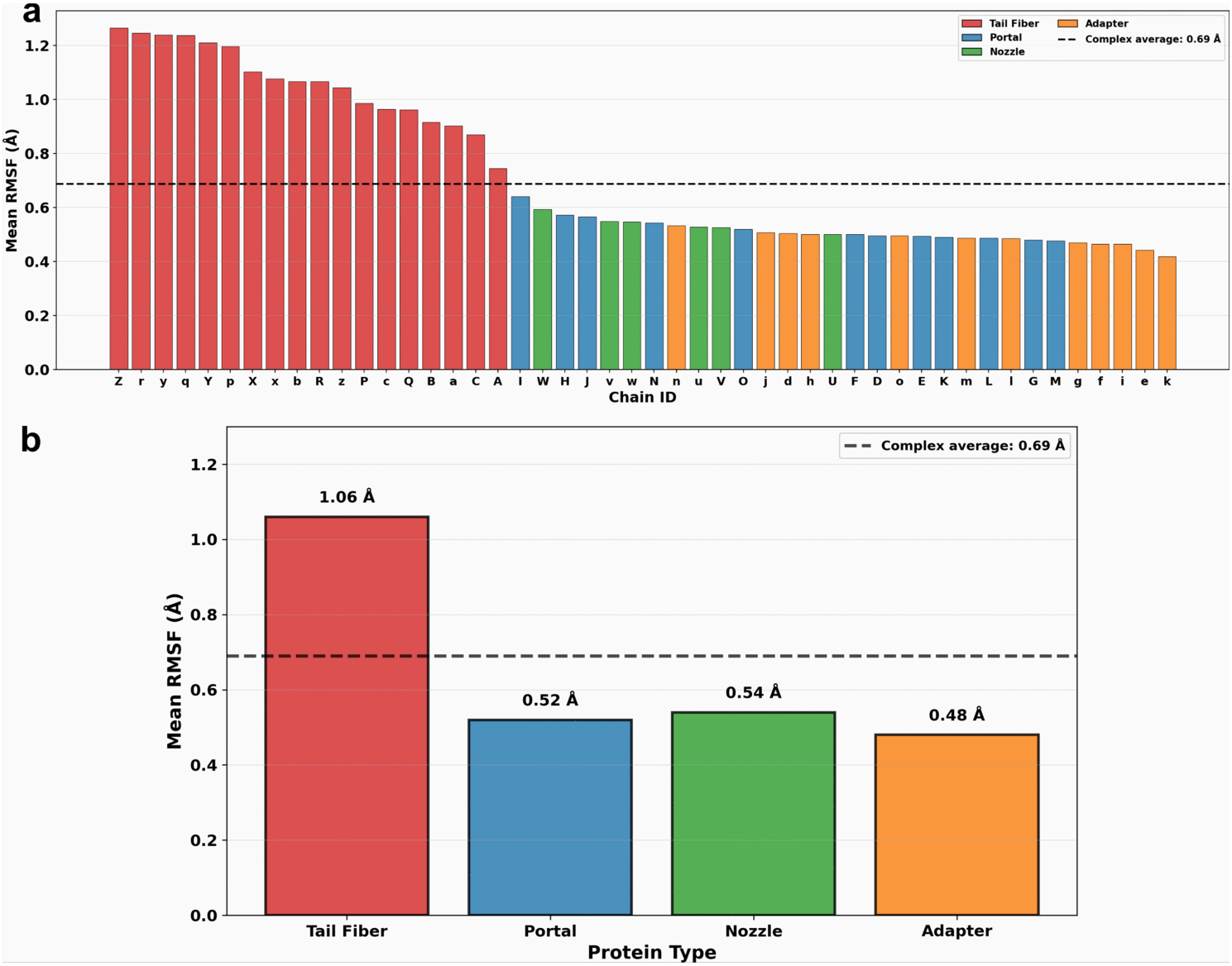


Figure 4

Per-chain and protein-type flexibility analysis of the cyanopodophage A4 portal-tail complex. a Mean RMSF for all 48 individual chains ranked by flexibility (50-250 ns). Chains are colored by protein type: tail fiber (red), portal (blue), nozzle (green), and adapter (orange). **b** Mean RMSF averaged across all chains of each protein type. Dashed black line indicates complex-wide average (0.69 Å).

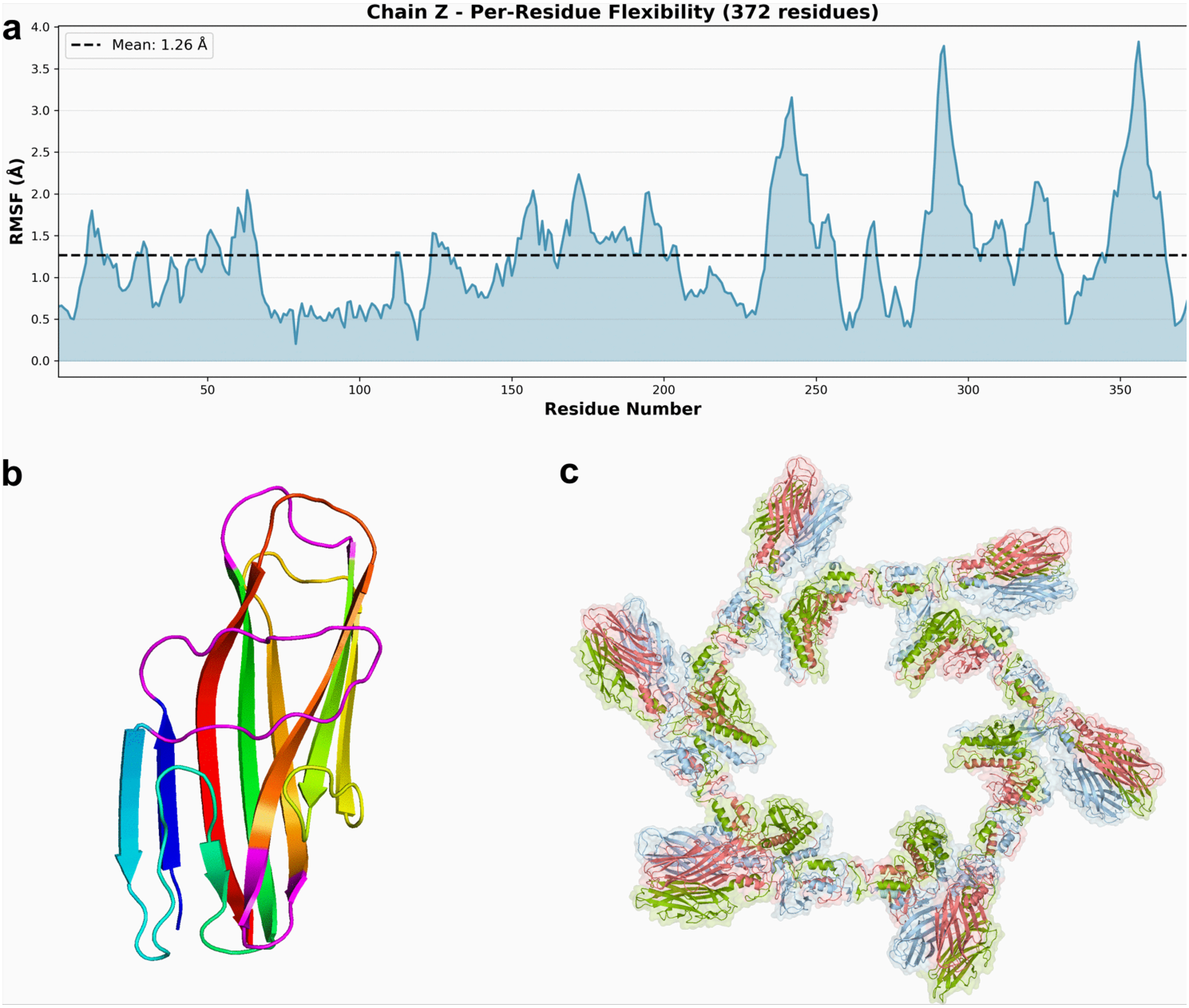


Figure 5

Residue-level flexibility of tail fiber chain Z and structural positions of the 18 tail fiber chains. a Per-residue RMSF for chain Z over the 50-200 ns production trajectory. The dashed line indicates the chain-wide mean (1.26 Å). **b** Structure of the C-terminal beta-sandwich domain (cell wall binding domain, residues 231–372) of chain Z (PDB: 9K09), with the most flexible regions highlighted in magenta. The remaining residues of the domain are shown in rainbow coloring (N-terminus, blue → C-terminus, red). **c** The 18 tail fiber chains of the portal-tail complex (PDB: 9K09, cartoon representation), color-coded by their structural position relative to the portal axis: proximal-to-capsid (red), interior (blue), and distal (green).

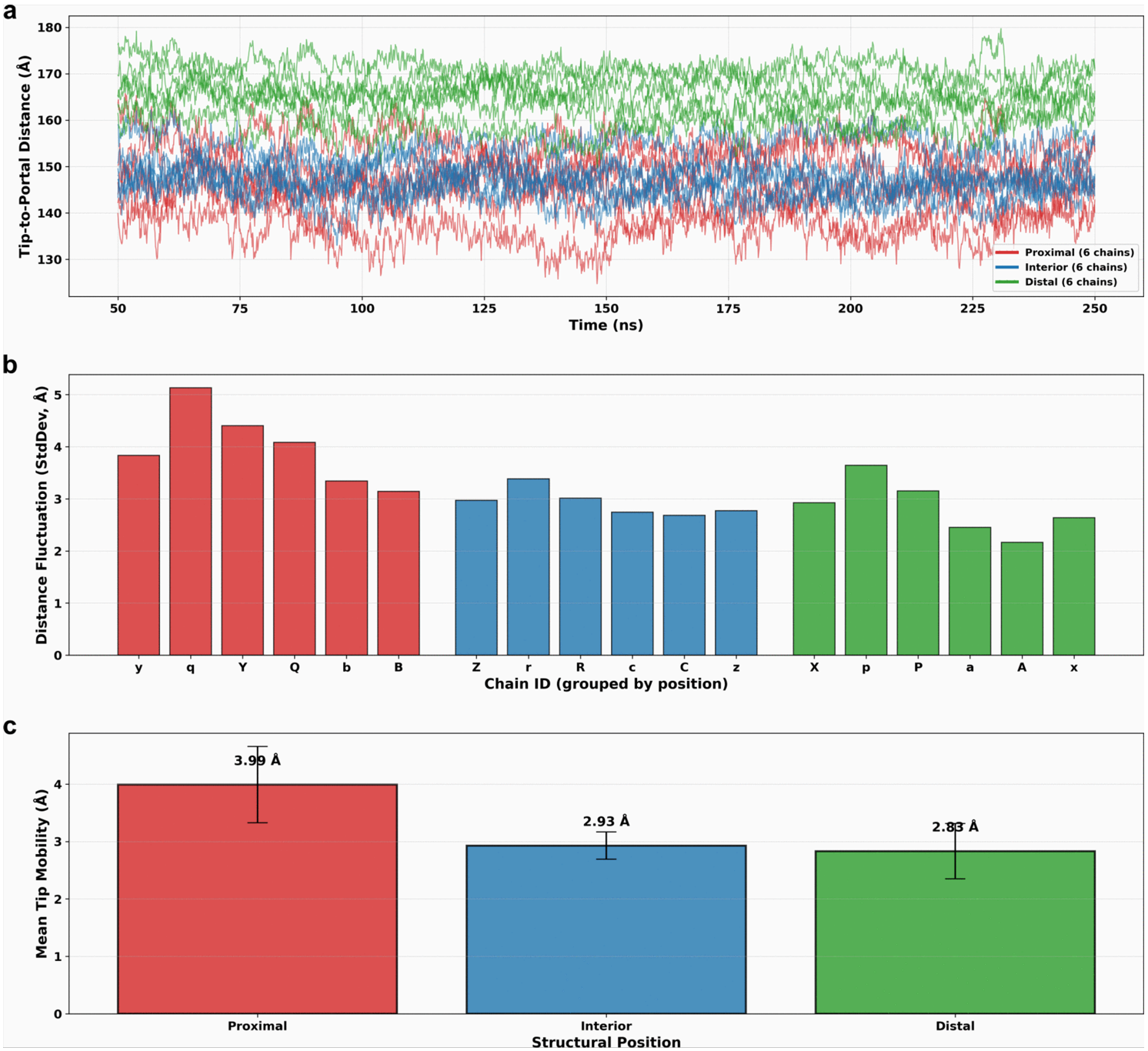


Figure 6

Tail fiber tip dynamics analysis. a Distance from each of the 18 tail fiber tips to the center of mass of the portal over the 50-250 ns production window. Chains are colored by structural position within the trimer: proximal-to-capsid (red, n = 6), interior (blue, n = 6), and distal (green, n = 6). **b** Standard deviation of tip-to-portal distance for each chain over the production window, grouped by structural position. **c** Mean standard deviation for each structural group. Error bars indicate the standard deviation across the six chains within each group.