Letter

Role of genome topology in the stability of viral capsids

James Daniel Farrell [1,2,*] Jure Dobnikar [1,2,3,†] and Rudolf Podgornik [2,4,1,5,‡]

[1]*CAS Key Laboratory of Soft Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China*
[2]*School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*
[3]*Songshan Lake Materials Laboratory, Dongguan, Guangdong 523808, China*
[4]*Kavli Institute for Theoretical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*
[5]*Wenzhou Institute of the University of Chinese Academy of Sciences, Wenzhou, Zhejiang 325000, China*

(Received 1 August 2022; accepted 2 February 2023; published 21 March 2023)

We explore how the stability of RNA viruses depends on genome topology and interactions between RNA and the capsid proteins. RNA is modeled as a branched polymer with 12 attractive sites (packaging signals) that can form bonds with 12 icosahedrally distributed capsid sites. The genome topology is encoded as a graph by mapping pairs of adjacent packaging signals to edges. We perform replica exchange molecular dynamics simulations and evaluate the osmotic pressure of all unique branched topologies of encapsulated RNA. We find that virion stability depends in a complex fashion on both genome topology and degree of confinement, and predict that MS2 bacteriophage should prefer a more linear genome topology.

DOI: 10.1103/PhysRevResearch.5.L012040

Introduction. The assembly of virions, i.e., viral capsids containing genetic material, is a key process in the reproductive cycle of viruses [1]. While bacteriophages and double stranded DNA (dsDNA) viruses use molecular motors [2] to pack the genome into pre-assembled capsids, single-stranded (ssRNA or ssDNA) viruses package their genomes concurrently with capsid assembly [3]. Since the N-terminal tails of capsid proteins (CPs) are positively charged, the negatively charged genome molecules provide a nonspecific electrostatic driving force [4–6] for the *co-assembly* process [7]. Additionally, partially hybridized encapsidated ssRNA features single-stranded motifs (∼20 base pair long stem loops) [8] that can form specific hydrogen bonds with CPs, triggering conformational changes [9] promoting assembly [10,11]. We refer to these motifs as *packaging signals* (PSs) binding to *capsid sites* (CSs). It is assumed that during co-assembly, RNA acts as a template recruiting CPs and promoting the assembly of shells with CS positions encoded as the vertices of a polyhedral graph with icosahedral symmetry [12].

The optimal RNA conformation and the connection between its topology and capsid symmetry [4,13], are poorly understood. The *Hamiltonian path hypothesis* [14] conjectures a minimum free-energy conformation where each dsRNA segment connects two adjacent CSs corresponding to an edge in the underlying polyhedral graph (see Fig. 1). This assumption is compatible with crystallographic and cryo-EM observations [15–17]. Several other experiments demonstrate that virion self-assembly relies on specific CS-PS interactions [10,18–20] that enable differentiation between cognate and nonviral RNA.

Here, we use replica exchange molecular dynamics simulations to explore the effect of genome topology on the thermodynamics of assembled icosahedral virions. We set up a coarse-grained model of RNA with 12 PSs, identify all unique topologies, and evaluate their osmotic pressure on the capsid walls. Our results show that the complex interplay of interactions, topology, and capsid geometry gives rise to rich behavior with regimes dominated by entropy, bond formation, and elasticity. The optimal chain topology changes from linear in large capsids (forming Hamiltonian paths) to highly branched in smaller capsids (forming bridged structures).

Topology. The genome topology can be encoded as a graph by mapping adjacent PSs to edges. Assuming a one-to-one CS:PS ratio, the spatial distribution of CSs generates a hard constraint: PSs that bind to any pair of CSs must be at least as "far apart" along the genome as the CSs are in space. The Hamiltonian path hypothesis (HPH) offers a further, softer constraint: for maximally efficient self-assembly, adjacent PSs must bind to adjacent CSs. In such a configuration, each chain segment bridges a pair of CSs that correspond to an edge in the underlying polyhedral graph (first-order bridge). Longer segments can form higher-order bridges between nonadjacent CSs (an nth-order bridge connects CSs n edges apart). With 12 PSs and 12 CSs, these constraints limit the genome topology to one of the 434 symmetry-unique spanning trees of the icosahedral graph; those are the topologies we explore here. They include a linear chain which can bind to CSs as a Hamiltonian path, and branched chains which can adopt

*farrelljd@iphy.ac.cn
†jd489@cam.ac.uk
‡podgornikrudolf@ucas.ac.cn; Also at Department of Physics, Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, SI-1000 Ljubljana, Slovenia.

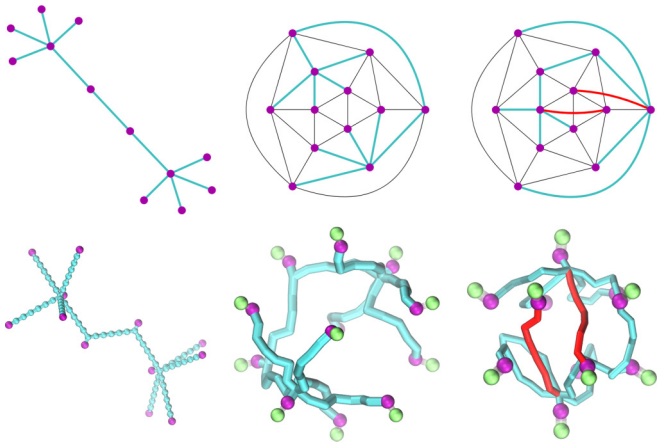

FIG. 1. Illustrations of the ssRNA model for the *bolus* topology. The illustration shows the mapping of the chain and the capsid to a tree/spanning tree (upper) and the corresponding snapshots from numerical simulations at two different capsid sizes (lower). Cyan, purple, and green spheres represent type A, B, and C beads, respectively (see text for their definitions). The bridges are shown as links in cyan (first-order bridges) and red (second-order bridges). Apart from the CSs, the spherical capsid is not shown.

non-Hamiltonian configurations (see Fig. 1). It is important to notice that "linear" does not imply unhybridized; we assume that the degree of partial hybridization of RNA does not vary with branchedness, achieved in our model by keeping the bending modulus Γ equal across all topologies.

Pressure. Our goal is to determine the effect of genome topology on capsid stability. Instead of evaluating the free-energy difference between a formed capsid and a reference state, we use osmotic pressure exerted by the genome on the capsid walls as a proxy for thermodynamic stability. This approach avoids the difficult problem of defining a proper reference state. In previous studies of the role of genome secondary structure on the stability of capsid [21], the conclusions obtained through the osmotic pressure are completely corroborated by the more detailed free-energy calculations. The pressure is computed as a sum of two terms: $P = P_{\text{surf}} + P_{\text{CS}}$. P_{surf} is a sum of nonspecific repulsive capsid-chain forces, while P_{CS} includes (attractive and repulsive) normal components of the forces on the CSs.

Topological indices. Theoretical studies of encapsidated RNA with nonspecific electrostatics [22] suggest a negative correlation between the number of annealed branch points and the capsid osmotic pressure [21]. The structure-property relationship is more complex with specific CS-PS interactions [23]. The branchedness of RNA genomes is often quantified with the ensemble-averaged maximum ladder distance ($\langle \text{MLD} \rangle$), i.e., a statistical average over the diameters of the graphs of RNA secondary structures [24–26]. For the small graphs considered here, MLD is a relatively coarse quantifier; to resolve the finer structure, we invoke the Wiener index W [27] and the number of angles, n_θ,

$$W = \sum_{u \neq v} d_{uv}\,; \quad n_\theta = \frac{1}{2} \sum_{u} \deg(u)\,[\deg(u) + 1], \qquad (1)$$

TABLE I. Three representative topologies along with values of topological indices. The vertices are colored by their degree: dark blue for deg = 1, light blue for 2, gray for 3, and red for deg = 5.

	Linear	Bolus	Wiener
Topology			
MLD	11	5	5
W	286	194	176
n_θ	10	22	18

where d_{uv} is the graph distance between vertices u and v, and $\deg(u)$ is the degree of (number of edges made by) vertex u. The more compact the topology, the smaller W and MLD, and the larger n_θ, as demonstrated in Table I. The linear chain takes on extreme values for all three indices, while $W^{\min} = 176$ ("Wiener") and $n_\theta^{\max} = 22$ ("bolus") are distinct topologies, which cannot be distinguished by MLD.

The indices W and n_θ probe different properties: n_θ is sensitive to local structure, e.g., nearest-neighbor PS distance, while W correlates with global properties, e.g., radius of gyration R_g of the chain [28]. We thus expect W to correlate with the behavior of a weakly confined genome at high temperatures where the polymer degrees of freedom are efficiently explored, while n_θ should work better in smaller capsids where long-range order is precluded by strong confinement and bond formation.

Simulations. In what follows, we use dimensionless units, measuring lengths in units of σ, energy in ϵ, temperature in ϵ/k_B, and pressure in ϵ/σ^3. We approximate an encapsidated genome by mapping a tree to a branched chain of soft, repulsive beads of diameter σ. The unit length is set to $\sigma = 2$ nm, as in similar models of DNA [29]. This choice ensures that the persistence length of the model at $\Gamma = 1$ and $T = 1$ corresponds to that of unfolded RNA at room temperature [30]. A single coarse-grained bead contains between 6 and 12 bases, depending on the degree of local hybridization. Vertices in the tree of the genome graph are mapped to a single bead (type A), and each edge between the vertices is represented by five A beads. Another bead (type B, representing PSs) is connected to each of the 12 vertex beads, giving a total of $N = (11 \times 5 + 12)A + 12B = 79$ beads (i.e., 500–1000 bases). Beads are bonded by two-body stretching $V_b = 16\,\Gamma(r-1)^2$ and three-body bending potentials $V_\theta = \Gamma(1 + \cos\theta)$, where r is the distance between two consecutive beads, and θ is the angle between consecutive bond vectors (the ratio 16 follows the exact result for elastic rods [31]). Additionally, *A*–*A* and *A*–*B* pairs interact via a Weeks-Chandler-Anderson (WCA) repulsion with strength ϵ. Beads are confined to a sphere of radius R by the same WCA repulsion. Twelve additional beads (type C, representing CSs) icosahedrally decorate the sphere interior, interacting with B beads via a 12-6 Lennard-Jones potential with well depth ϵ_{BC}. Finally, *B*–*B* pairs repel via a stronger WCA than the other beads (strength ϵ_{BC} and effective diameter 2σ) to prevent multiple B binding to one C, reflecting the specific nature of PS-CS binding in viruses [9]. The real nature of the interactions is of course electrostatic, involv-

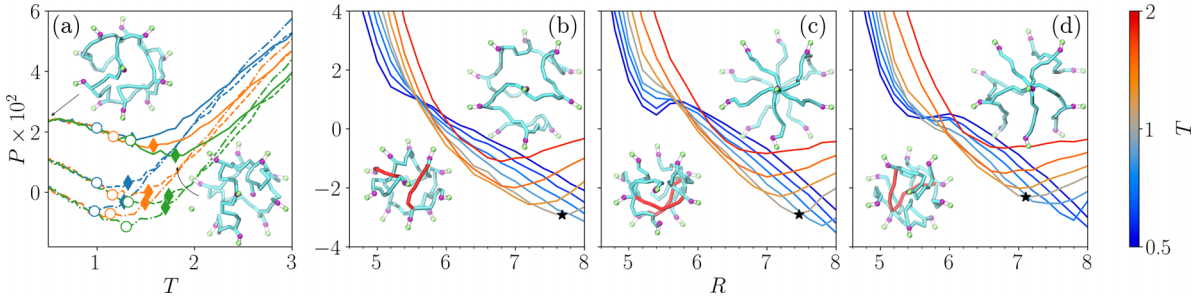

FIG. 2. (a) Pressure as a function of temperature for a linear genome in a capsid $R = 6.2$. Blue, orange, and green lines correspond to $\epsilon_{BC} = 12, 14, 16$; solid, dashed, and dash-dotted lines correspond to $\Gamma = 1, 2, 4$. Symbols mark where 97% and 10/12 bonds are formed on average (empty circles and full diamonds, respectively). Typical snapshots are shown for the harmonic and attractive regimes. Results for other topologies and capsid sizes are summarized in Fig. S1 in the Supplemental Material [46]. (b)–(d) Pressure against radius at select temperatures for the (b) linear, (c) bolus, and (d) Wiener topologies, with snapshots at $R = 5.4$ and 7.2 ($T = 0.5$). CSs are shown as green spheres, PSs as purple spheres, and chain segments as cyan or red tubes, indicating a first- or second-order bridge, respectively. Black stars indicate the radius of minimum pressure at $T = 1$.

ing strong, nonlinear many-body effects and complex charge regulation mechanisms operating between protein N-tails and RNA, which are computationally expensive to model. Here, we implemented a much simpler pairwise-additive WCA interaction to model repulsive interactions. The WCA model captures the general physics of the problem and is computationally cheap to compute, allowing for a comprehensive study of the effect of the chain topology on thermodynamic stability [32].

The packing density of our model genome with fixed length is controlled by R alone. The accessible interval is $4 \lesssim R \lesssim 8$ (at smaller radii, simulations are frustrated by high density and chain rigidity, while at larger radii, sampling of fully bound configurations is poor). Therefore, we can access packing fractions roughly between 0.05 to 0.5 bases/nm^3, which falls within the range of known RNA viruses (0.05–1 bases/nm^3 [34]).

We performed NVT replica-exchange molecular dynamics simulations [35,36] using the LAMMPS package [37], simulating 24 replicas with temperature T regulated by a Langevin thermostat for up to 2×10^9 steps, attempting exchanges between replicas every 10^4 steps and dumping snapshots every 10^6 steps. We chose a geometric temperature distribution $T \in [0.5, 2]$, which produced exchange probabilities between 46% and 76%. We employ a variable time step $\Delta\tau$, which is dynamically determined such that no bead moves by more than $0.018\sigma/\Delta\tau$, leading to $\Delta\tau = 0.005\tau$ at $T = 1$. To dampen statistical noise, we smooth the temperature-series pressure data using a Savitsky-Golay filter [38], whereafter we measure the linear correlation between pressure and topological indices at fixed R, T via the Pearson correlation coefficient r. Initial configurations were generated, analyses performed, and graphs drawn with the aid of the networkx [39], pynauty, numpy [40], scipy [41], pandas [42,43], and matplotlib [44] libraries. Simulation snapshots were drawn in VMD [45].

Results. We have performed extensive simulations evaluating the pressure P by varying genome topology, chain rigidity Γ, attraction strength ϵ_{BC}, temperature T, and capsid radius R. Some generic observations are summarized in Fig. 2(a), where the pressure exerted on the capsid with radius $R = 6.2$ by a linear genome is plotted as a function of temperature for various ϵ_{BC} and Γ.

We observe qualitatively distinct regimes. At low-enough temperatures, all 12 bonds are formed and the PS-CS pairs behave as harmonic springs. In this *elastic regime*, the total confining pressure is independent of ϵ_{BC}, scales with Γ, and decreases linearly with T. As T increases, the individual contribution of each PS-CS pair to $P_{\rm CS}$ increases in magnitude due to thermal fluctuations, but the PS-CS bonds also start breaking, reducing the number of these contributions; as a result, the pressure develops a minimum as a function of T. The temperature of the pressure minimum varies nonmonotonically with radius, peaking at around $R = 6.2$ for $\epsilon_{BC} = 16$, $\Gamma = 1$ (see Fig. S1 in the Supplemental Material [46]). Beyond the minimum, in the *attractive regime*, the pressure is dominated by the CS-PS attraction ϵ_{BC}, and is proportional to the number of formed bonds [47]. This behavior is universal at large-enough R regardless of the genome topology, but at small R, second-nearest-neighbor CSs are close enough to form second-order bridges, which are entropically disfavored but lower both the elastic energy and $P_{\rm CS}$ with respect to first-order ones; as a result, the pressure continues to decrease even when all PSs are bound, as second-order bridges replace first-order ones.

At a fixed T, P exhibits a minimum as a function of radius, suggesting a preferred capsid size [see Figs. 2(b)–2(d)]. Reducing the radius below the minimum increases P: at high T, it increases $P_{\rm surf}$, and at low T, it decreases $P_{\rm CS}$ by removing tension from the bridges. Similarly, increasing R also increases P by lowering the entropy of bound configurations and therefore the number of bonds. While these trends are qualitatively preserved across genome topologies, we observe that pressure also depends on topology in a complex manner, which varies with both T and R.

In the attractive regime at large R, less-branched genomes typically exert more-negative pressures on their capsids [see Fig. 3(a)]. This effect is most pronounced when an average of 10/12 bonds has formed and is strongly associated with both W and n_θ. To understand this, we distinguish between vertices of a low degree ("peripheral") and those of a higher

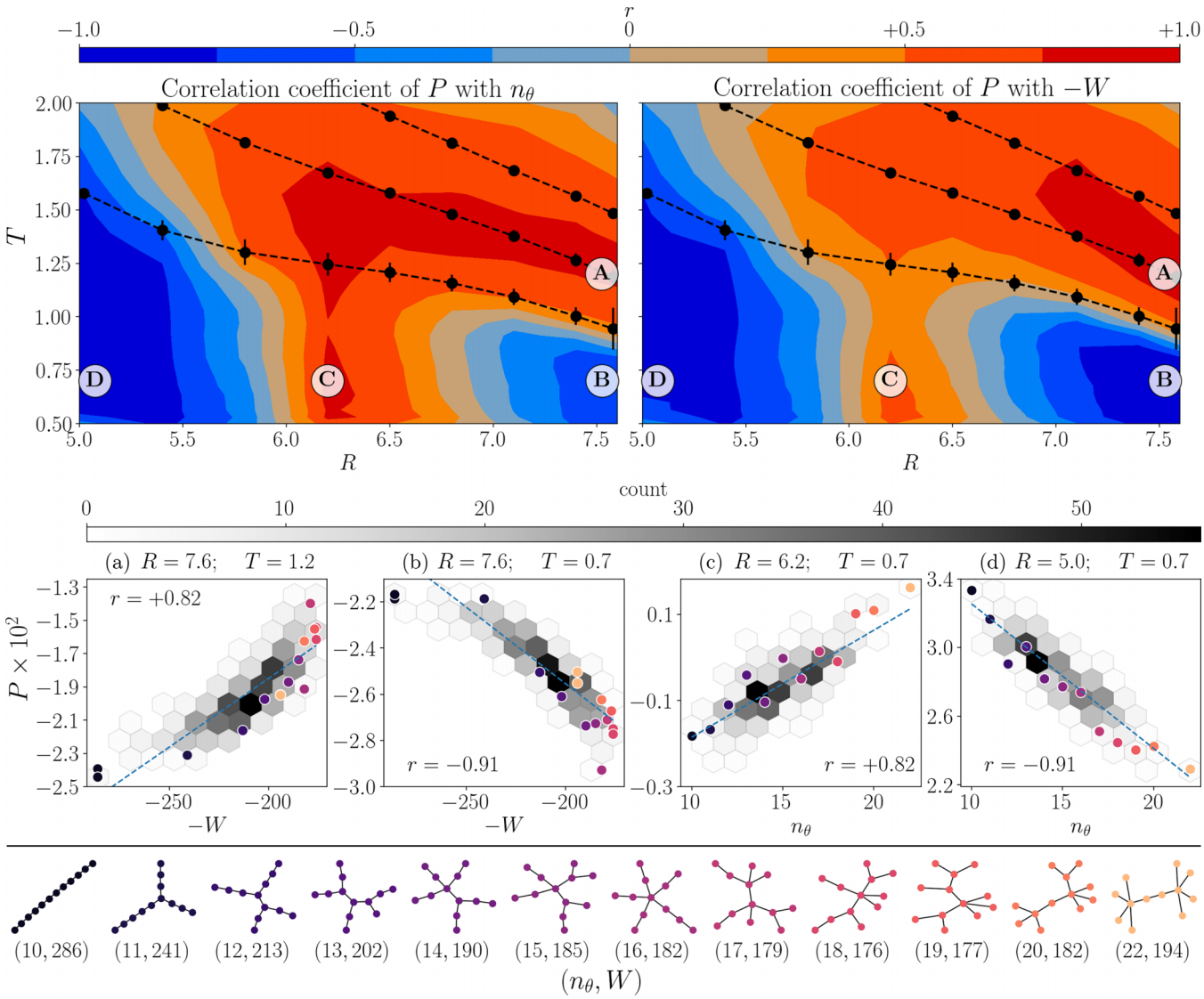

FIG. 3. The relationship between pressure and various measures of branchedness depends sensitively on both temperature and capsid radius. Top panel: Correlation coefficient r between P and n_θ (left) and $-W$ (right) as a function of T and R. From top to bottom, the dotted lines are drawn where, on average, 8 of 12, 10 of 12, and 97% of PSs are bound to CSs. Lower panel: Least-squares fits of P to n_θ or $-W$ at select (R, T). Blue dashed lines are lines of best fit. At large R, P (A) decreases with branchedness at high T, but (B) increases with branchedness at low T. For all T, P increases with branchedness (n_θ) at intermediate R [see, e.g., (C)], but decreases with branchedness at small R [e.g., (D)].

degree ("central"). Peripheral PSs have a larger propensity to form bonds than central ones (see Fig. S2 in the Supplemental Material [46]): at $R = 7.4$, degree-one PSs are bound 95% of the time at $T \approx 1.32$, while the same binding is achieved by degree-five PSs at $T \approx 0.86$.

Conversely, in the harmonic regime, pressure is negatively correlated with branchedness; it is most strongly correlated with W, suggesting a global topological feature [see Fig. 3(b)]. More-branched topologies are more compact (smaller $\langle R_g \rangle$) and less compressible [smaller $\mathrm{Var}(R_g)$] (see Fig. S3 in the Supplemental Material [46]); as such, fully bound configurations exert a greater pulling force on the capsid. This is reflected in the preference of more-(Wiener)-branched genomes for more compact enclosures [Figs. 2(b)–2(d)]: at $T = 1$, the linear, bolus, and Wiener topologies prefer capsids with radii 7.7, 7.5, and 7.1, respectively.

Reducing the capsid radius in the harmonic regime both reduces the magnitude of the capsid pressure and brings about a weak preference for less-branched topologies, most pronounced at $R = 6.2$ [see Fig. 3(c)]. In this more compact regime, being less compressible elevates P_{surf}. Additionally, since less-branched topologies exhibit a smaller mean neighbor distance $\langle d_{ij} \rangle$, their bridges are held under greater tension, producing a lower P_{CS}.

Second-order bridges become abundant at small R and low T [48]. They lower P by localizing chain segments away from the capsid surface and by orienting the CS-PS

bonds towards the capsid center, increasing the nonlocal elastic bridging attraction. More-branched topologies more often form second-order bridges (Fig. S4 in the Supplemental Material [46]; such topologies are weakly preferred at $R = 5.8$ and strongly preferred at $R = 5.4$ and $R = 5.0$ [see Fig. 3(d)]). Moreover, the pressure exerted by highly branched topologies on the capsid walls develops a second, low-T minimum as a function of R [see Figs. 2(c) and 2(d)].

Discussion and conclusions. Extensive simulations revealed the connection between the topology of RNA genomes and the structural stability of co-assembled virions as quantified by their osmotic pressure. We identified the topological indices describing the branchedness of RNA that best correlate with the computed pressure (W and n_θ). All 434 symmetry-unique topologies can therefore be classified—and their behavior in capsid confinement predicted—by evaluating these quantifiers.

Over a wide parameter range, pressure develops a minimum as a function of R. The optimal R depends on genome topology, suggesting branchedness as a selection criterion for capsid size. Experimental data with which we can compare our predictions are not abundant. Cryo-EM experiments on MS2 bacteriophage [49] identify 59 stem loops (PSs) on the encapsidated RNA, 44 of which (around 75%) are co-localized with CP_2 RNA binding sites (CSs). The radius and genome length of the MS2 virus are ≈25 nm and 3569 bases [50], corresponding to packing density (≈0.06 bases/nm^3), i.e., $R = 6.3$. From Fig. 3, we estimate our model predicts a fraction of the formed bonds comparable to the experimental value at $T \approx 2$. Hence, our prediction is that MS2 prefers a linear genome topology, which supports the Hamiltonian path hypothesis [14].

Highly branched topologies produce a second minimum under strong confinement, where RNA forms bridging conformations. The fact that branched RNA can enable bistable virion assembly is interesting, and we speculate that the existence of multiple packaging states could play a role in multipartite RNA viruses [51], enabling maintenance of the genomic integrity and virus identity in viruses with segmented genomes [52].

Although not addressed here, it is easy to imagine that the topology of the genome will also have a strong effect on the co-assembly kinetics [13,53–63]; topologies that form more stable virions may also have frustrated energy landscapes with many kinetic traps, impeding self-assembly. Furthermore, the most stable virion need not be the biologically most fit since both assembly and disassembly are essential parts of the viral life cycle. Thus, it would be extremely interesting to compare our results with experiments and kinetic assembly simulations in the future, to complete the connection between virus self-assembly dynamics and thermodynamics.

Acknowledgments. We thank Reidun Twarock for communications about PS-mediated genome packing. We acknowledge funding from the Key Project No. 12034019 of the National Natural Science Foundation of China, from the Chinese National Science Foundation (Grants No. 11874398 and No. 12034019), Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB33000000), and by the international collaboration grant from the K. C. Wong Educational Foundation.

[1] R. Zandi, B. Dragnea, A. Travesset, and R. Podgornik, On virus growth and form, Phys. Rep. **847**, 1 (2020).

[2] W. H. Roos, R. Bruinsma, and G. J. L. Wuite, Physical virology, Nat. Phys. **6**, 733 (2010).

[3] M. Comas-Garcia, Packaging of genomic RNA in positive-sense single-stranded RNA viruses: A complex story, Viruses **11**, 253 (2019).

[4] R. Twarock, R. J. Bingham, E. C. Dykeman, and P. G. Stockley, A modelling paradigm for RNA virus assembly, Curr. Opin. Virol. **31**, 74 (2018).

[5] H.-X. Zhou and X. Pang, Electrostatic interactions in protein structure, folding, binding, and condensation, Chem. Rev. **118**, 1691 (2018).

[6] V. A. Belyi and M. Muthukumar, Electrostatic origin of the genome packing in viruses, Proc. Natl. Acad. Sci. USA **103**, 17174 (2006).

[7] A. Klug, The tobacco mosaic virus particle: Structure and assembly, Philos. Trans. R. Soc. London B **354**, 531 (1999).

[8] R. Twarock and P. G. Stockley, RNA-mediated virus assembly: Mechanisms and consequences for viral evolution and therapy, Annu. Rev. Biophys. **48**, 495 (2019).

[9] L. Poudel, R. Twarock, N. F. Steinmetz, R. Podgornik, and W.-Y. Ching, Impact of hydrogen bonding in the binding site between capsid protein and MS2 bacteriophage ssRNA, J. Phys. Chem. B **121**, 6321 (2017).

[10] P. G. Stockley, R. Twarock, S. E. Bakker, A. M. Barker, A. Borodavka, E. Dykeman, R. J. Ford, A. R. Pearson, S. E. Phillips, and N. A. Ranson, Packaging signals in single-stranded RNA viruses: Nature's alternative to a purely electrostatic assembly mechanism., J. Biol. Phys. **39**, 277 (2013).

[11] A. Routh, T. Domitrovic, and J. Johnson, Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus., Proc. Natl. Acad. Sci. USA **109**, 1907 (2012).

[12] R. Twarock, G. Leonov, and P. G. Stockley, Hamiltonian path analysis of viral genomes., Nat. Commun. **9**, 2021 (2018).

[13] I. Mizrahi, R. Bruinsma, and J. Rudnick, Spanning tree model and the assembly kinetics of RNA viruses, Phys. Rev. E, **106**, 044405 (2022).

[14] J. Rudnick and R. Bruinsma, Icosahedral packing of RNA viral genomes, Phys. Rev. Lett. **94**, 038101 (2005).

[15] S. J. Schroeder, Probing viral genomic structure: Alternative viewpoints and alternative structures for satellite tobacco mosaic virus RNA, Biochemistry **53**, 6728 (2014).

[16] S. B. Larson, J. S. Day, and A. McPherson, *Satellite tobacco mosaic virus* refined to 1.4Å resolution, Acta Crystallogr. Sect. D **70**, 2316 (2014).

[17] D. Luque and J. R. Castón, Cryo-electron microscopy for the study of virus assembly, Nat. Chem. Biol. **16**, 231 (2020).

[18] N. Patel, E. C. Dykeman, R. H. A. Coutts, G. P. Lomonossoff, D. J. Rowlands, S. E. V. Phillips, N. Ranson, R. Twarock,

R. Tuma, and P. G. Stockley, Revealing the density of encoded functions in a viral RNA., Proc. Natl. Acad. Sci. USA **112**, 2227 (2015).

[19] E. C. Dykeman, P. G. Stockley, and R. Twarock, Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy., Proc. Natl. Acad. Sci. USA **111**, 5361 (2014).

[20] M. Comas-Garcia, T. Kroupa, S. A. Datta, D. P. Harvin, W.-S. Hu, and A. Rein, Efficient support of virus-like particle assembly by the HIV-1 packaging signal., eLife **7**, e38438 (2018).

[21] G. Erdemci-Tandogan, J. Wagner, P. van der Schoot, R. Podgornik, and R. Zandi, Effects of RNA branching on the electrostatic stabilization of viruses, Phys. Rev. E **94**, 022408 (2016).

[22] L. Marichal, L. Gargowitsch, R. L. Rubim, C. Sizun, K. Kra, S. Bressanelli, Y. Dong, S. Panahandeh, R. Zandi, and G. Tresset, Relationships between RNA topology and nucleocapsid structure in a model icosahedral virus, Biophys. J. **120**, 3925 (2021).

[23] C. Huang, R. Podgornik, and X. Man, Selective adsorption of confined polymers: Self-consistent field theory studies, Macromolecules **54**, 9602 (2021).

[24] A. M. Yoffe, P. Prinsen, A. Gopal, C. M. Knobler, W. M. Gelbart, and A. Ben-Shaul, Predicting the sizes of large RNA molecules, Proc. Natl. Acad. Sci. USA **105**, 16153 (2008).

[25] L. Tubiana, A. L. Božič, C. Micheletti, and R. Podgornik, Synonymous mutations reduce genome compactness in icosahedral ssRNA viruses, Biophys. J. **108**, 194 (2015).

[26] R. F. Bruinsma, G. J. L. Wuite, and W. H. Ross, Physics of viral dynamics, Nat. Rev. Phys. **3**, 76 (2021).

[27] H. Wiener, Structural determination of paraffin boiling points, J. Am. Chem. Soc. **69**, 17 (1947), pMID: 20291038.

[28] P. F. Sheridan, D. B. Adolf, A. V. Lyulin, I. Neelov, and G. R. Davies, Computer simulations of hyperbranched polymers: The influence of the Wiener index on the intrinsic viscosity and radius of gyration, J. Chem. Phys. **117**, 7802 (2002).

[29] T. Curk, J. D. Farrell, J. Dobnikar, and R. Podgornik, Spontaneous domain formation in spherically confined elastic filaments, Phys. Rev. Lett. **123**, 047801 (2019).

[30] G. Caliskan, C. Hyeon, U. Perez-Salas, R. M. Briber, S. A. Woodson, and D. Thirumalai, Persistence Length Changes Dramatically as RNA Folds, Phys. Rev. Lett. **95**, 268303 (2005).

[31] R. Podgornik, P. L. Hansen, and V. A. Parsegian, Elastic moduli renormalization in self-interacting stretchable polyelectrolytes, J. Chem. Phys. **113**, 9343 (2000).

[32] The screened Coulomb (Yukawa) interactions do not capture the details of complex electrostatic interactions in our system, and they are computationally more demanding than WCA since they have a longer range. In fact, based on the Onsager argument for charged cylinders [33], we anticipate that the results of a Yukawa model would qualitatively correspond to WCA with thicker chains.

[33] P. Van Der Schoot, in *Molecular Theory of Nematic (and other) Liquid Crystals*, SpringerBriefs in Physics (Springer, Cham, 2022), Chap. 1, pp. 1–14.

[34] J. Cui, T. E. Schlub, and E. C. Holmes, An allometric relationship between the genome length and virion volume of viruses, J. Virol. **88**, 6403 (2014).

[35] U. H. Hansmann, Parallel tempering algorithm for conformational studies of biological molecules, Chem. Phys. Lett. **281**, 140 (1997).

[36] Y. Sugita and Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, Chem. Phys. Lett. **314**, 141 (1999).

[37] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, J. Comput. Phys. **117**, 1 (1995).

[38] R. W. Schafer, What is a Savitzky-Golay filter?, IEEE Signal Process Mag. **28**, 111 (2011).

[39] A. Hagberg, P. Swart, and D. S Chult, Exploring network structure, dynamics, and function using NetworkX, in *Proceedings of the 7th Python in Science conference (SciPy 2008)*, edited by G. Varoquaux, T. Vaught, and J. Millman (SciPy, 2008), pp. 11–15.

[40] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant *et al.*, Array programming with NumPy, Nature (London) **585**, 357 (2020).

[41] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey *et al.*, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nat. Methods **17**, 261 (2020).

[42] The Pandas Development Team, pandas-dev/pandas: Pandas (2020), doi: 10.5281/zenodo.3509134.

[43] Wes McKinney, Data structures for statistical computing in python, in *Proceedings of the 9th Python in Science Conference*, edited by S. van der Walt and J. Millman (SciPy, 2010), pp. 56–61.

[44] J. D. Hunter, Matplotlib: A 2D graphics environment, Comput. Sci. Eng. **9**, 90 (2007).

[45] W. Humphrey, A. Dalke, and K. Schulten, VMD – Visual Molecular Dynamics, J. Mol. Graphics **14**, 33 (1996).

[46] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.5.L012040 for pressure-temperature curves at different topologies and radii; binding temperatures of PSs as a function of degree and R; the relationships of R_g and d_{ij} with topological indices; and more detail about the frequency of bridges as a function of R and n_θ.

[47] At high T, most PSs unbind and the pressure asymptotically approaches the purely steric $\epsilon_{BC} \to 0$ limit, i.e., the *entropic regime*. This is expected at small packing densities and we have observed the onset of such regime in trial simulations of large capsids (data not shown). We have not explored it systematically as it is not biologically relevant [34].

[48] For the capsid sizes studied here, third-order bridges occur with negligible frequency.

[49] R. I. Koning, J. Gomez-Blanco, I. Akopjana, J. Vargas, A. Kazaks, K. Tars, J. M. Carazo, and A. J. Koster, Asymmetric cryo-em reconstruction of phage MS2 reveals genome structure in situ, Nat. Commun. **7**, 12524 (2016).

[50] W. Fiers, R. Contreras, F. Duerinck, G. Haegeman, D. Iserentant, J. Merregaert, W. Min Jou, F. Molemans, A. Raeymaekers, A. van den Berghe, G. Volckaert, and M. Ysebaert, Complete nucleotide sequence of bacteriophage MS2 RNA: Primary and secondary structure of the replicase gene, Nature (London) **260**, 500 (1976).

[51] A. Sicard, Y. Michalakis, S. Gutiérrez, and S. Blanc, The strange lifestyle of multipartite viruses, PLoS Pathog. **12**, e1005819 (2016).

[52] Y. Michalakis and S. Blanc, The curious strategy of multipartite viruses, Annu. Rev. Virol. **7**, 203 (2020).

[53] A. Kivenson and M. F. Hagan, Mechanisms of capsid assembly around a polymer, Biophys. J. **99**, 619 (2010).

[54] O. M. Elrad and M. F. Hagan, Encapsulation of a polymer by an icosahedral virus, Phys. Biol. **7**, 045003 (2010).

[55] J. Mahalik and M. Muthukumar, Langevin dynamics simulation of polymer-assisted virus-like assembly, J. Chem. Phys. **136**, 135101 (2012).

[56] J. D. Perlmutter and M. F. Hagan, The role of packaging sites in efficient and specific virus assembly, J. Mol. Biol. **427**, 2451 (2015).

[57] M. F. Hagan and R. Zandi, Recent advances in coarse-grained modeling of virus assembly, Curr. Opin. Virol. **18**, 36 (2016).

[58] C. Beren, L. L. Dreesens, K. N. Liu, C. M. Knobler, and W. M. Gelbart, The effect of RNA secondary structure on the self-assembly of viral capsids, Biophys. J. **113**, 339 (2017).

[59] M. Chevreuil, D. Law-Hine, J. Chen, S. Bressanelli, S. Combet, D. Constantin, J. Degrouard, J. Möller, M. Zeghal, and G. Tresset, Nonequilibrium self-assembly dynamics of icosahedral viral capsids packaging genome or polyelectrolyte, Nat. Commun. **9**, 3071 (2018).

[60] S. Li, G. Erdemci-Tandogan, P. van der Schoot, and R. Zandi, The effect of RNA stiffness on the self-assembly of virus particles, J. Phys.: Condens. Matter **30**, 044002 (2018).

[61] A. Zlotnick, R. Aldrich, J. M. Johnson, P. Ceres, and M. J. Young, Mechanism of capsid assembly for an icosahedral plant virus, Virology **277**, 450 (2000).

[62] L. Lavelle, M. Gingery, M. Phillips, W. Gelbart, C. Knobler, R. Cadena-Nava, J. Vega-Acosta, L. Pinedo-Torres, and J. Ruiz-Garcia, Phase diagram of self-assembled viral capsid protein polymorphs, J. Phys. Chem. B **113**, 3813 (2009).

[63] Y. Hu, R. Zandi, A. Anavitarte, C. M. Knobler, and W. M. Gelbart, Packaging of a polymer by a viral capsid: The interplay between polymer length and capsid size, Biophys. J. **94**, 1428 (2008).