

# Percolation Theory Reveals Biophysical Properties of Virus-like Particles

Nicholas E. Brunk\* and Reidun Twarock\*

Cite This: *ACS Nano* 2021, 15, 12988–12995

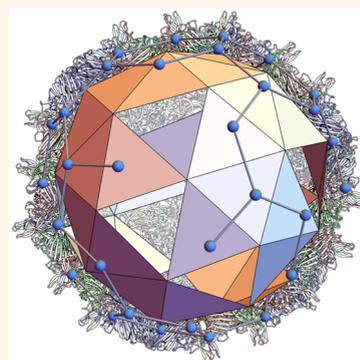
Read Online

ACCESS |

Metrics &amp; More

Article Recommendations

**ABSTRACT:** The viral protein containers that encapsulate a virus' genetic material are repurposed as virus-like particles in a host of nanotechnology applications, including cargo delivery, storage, catalysis, and vaccination. These viral architectures have evolved to sit on the knife's edge between stability, to provide adequate protection for their genetic cargoes, and instability, to enable their efficient and timely release in the host cell environment upon environmental cues. By introducing a percolation theory for viral capsids, we demonstrate that the geometric characteristics of a viral capsid in terms of its subunit layout and intersubunit interaction network are key for its disassembly behavior. A comparative analysis of all alternative homogeneously tiled capsid structures of the same stoichiometry identifies evolutionary drivers favoring specific viral geometries in nature and offers a guide for virus-like particle design in nanotechnology.



**KEYWORDS:** virus-like particle, virus disassembly, percolation theory, virus nanotechnology, generalized quasi-equivalence principle

## INTRODUCTION

Mining design features from nature is a cornerstone of virus nanotechnology. The molecular designs of the viral protein shells, called viral capsids, which surround and thus protect the viral genomes, are exploited as virus-like particles (VLPs) for cargo delivery and storage.<sup>1,2</sup> Their physical properties can be tuned to support the desired purposes,<sup>3–6</sup> for example, by modulating the charge of particles *via* peptide binding or through direct removal and replacement of protein subunits as a means of functionally decorating the capsid surface.<sup>7,8</sup> The net charge is also important for a variety of viral mechanisms, including assembly and the encapsidation of charged polymers.<sup>9–12</sup> Perforated capsids<sup>7,13</sup> are reminiscent of a breadboard with modular components. They offer the opportunity to fractionally refill the VLP surface in order to modulate its steric, elastic, and electrostatic properties.<sup>2,8</sup> Modular VLP technologies depend crucially on a VLP's resistance to fragmentation upon subunit removal prior to their replacement. Much like the shape of building blocks would affect the disassembly of a tower, as in the board game Jenga,<sup>14</sup> the shape and valency of the capsid building blocks (capsomers) affect the assembly and disassembly behavior of the virus. A capsid's geometric design and its implied subunit-bond network are therefore essential for its role in cargo encapsidation and release.

The number of distinct viral capsid designs in nature is limited. This is a consequence of the principle of genetic economy,<sup>15</sup> which stipulates that viral capsids are formed from

multiple copies of identical protein subunits synthesized from the same genomic fragment, thus reducing the overall length of the genomic sequence required to code for the capsid. As identical protein subunits form the same types of local interactions across the capsid surface, they self-assemble into capsids with a high degree of symmetry, and icosahedral viruses make up the vast majority of particle architectures in the virosphere. Caspar and Klug's quasi-equivalence theory models icosahedral viral capsid architecture *via* polyhedral blueprints, which are parametrized in terms of the triangulation number  $T$ ,<sup>16</sup> implying that capsids must be formed from precisely  $60T$  protein subunits. The cancer-causing papillomaviruses, however, form a notable exception, and prompted the generalization of their theory in Viral Tiling theory.<sup>17</sup> Recently, a comprehensive theory of viral capsid architecture has been introduced based on a generalized quasi-equivalence principle<sup>18</sup> that includes both Caspar–Klug theory and Viral Tiling theory as special cases.

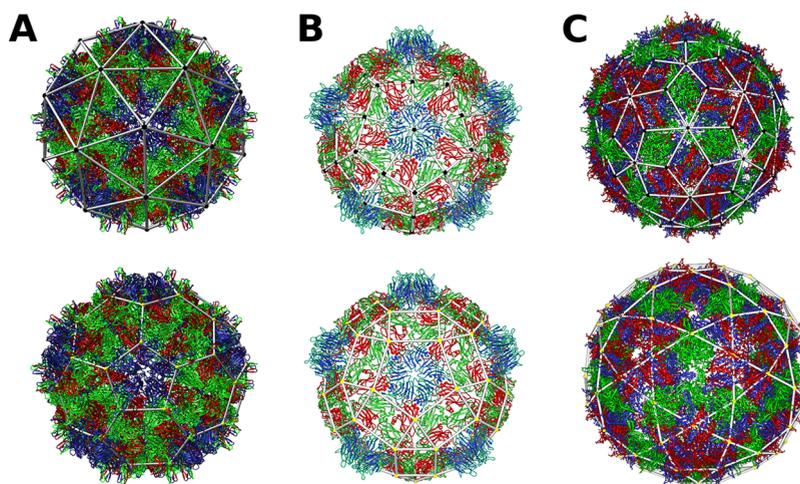
We analyze here the impact of distinct capsid designs in this classification on capsid disassembly. The importance of the

Received: March 3, 2021

Accepted: July 19, 2021

Published: July 23, 2021





**Figure 1.** Protein layouts and interaction networks for different capsid geometries. There are three distinct types of structural organization in the viral capsids that are classed as  $T = 3$  architectures in Caspar Klug theory: (A) the triangular tiling in which each subunit is trivalent, as is the case for Pariacoto virus (1f8v); (B) the kite-shaped tiling in which each subunit is tetravalent, as is the case for Tobacco Ring virus (1a6c); (C) the rhombic tiling in which each subunit is tetravalent, as is the case for bacteriophage MS2 (2ms2). Underneath are shown the dual tilings superimposed on the capsids, which encode the topology of the interaction network.

subunit-bond network topology for the resistance of the capsid to fragmentation and disassembly has been demonstrated previously for Hepatitis B virus (HBV) *in vitro*. In these experiments, subunits were removed from intact capsids, or from capsids in which chemical cross-linking prevented the removal of passivated subunits, using denaturant. Single particle mass spectrometry was used to interrogate the resulting particles, revealing a marked absence of any particles fewer than 90 protein dimers, in excellent agreement with the theoretically predicted fragmentation threshold.<sup>7,19</sup> However, the dependence of capsid disassembly on capsid geometry and topology has not been investigated before. We are closing this gap here by developing a generalized percolation theory for virus capsid disassembly. These fragmentation thresholds characterize the onset of dissociation of the capsid in terms of the numbers of subunits removed or of the intersubunit contacts broken. They are stricter versions of the traditional percolation thresholds, which characterize the subsequent, abrupt disappearance of all clusters of the order of system size. We investigate how the topologies of alternative capsid geometries with identical protein stoichiometries and their associated subunit-bond networks impact the VLP fragmentation as subunits are randomly removed or when bonds are randomly broken. In particular, the maximal number of subunits (or, respectively, bonds) that may be removed without, on average, inducing fragmentation of the remaining capsid shell (*i.e.*, the subunit and bond fragmentation thresholds) are computed, enabling a comparative analysis of the resilience to fragmentation of distinct capsid designs. Our results reveal the mechanistic pressures on viral evolution and provide a possible explanation for the abundance of specific viral capsid designs in nature. They also inform VLP design in bionanotechnology applications.

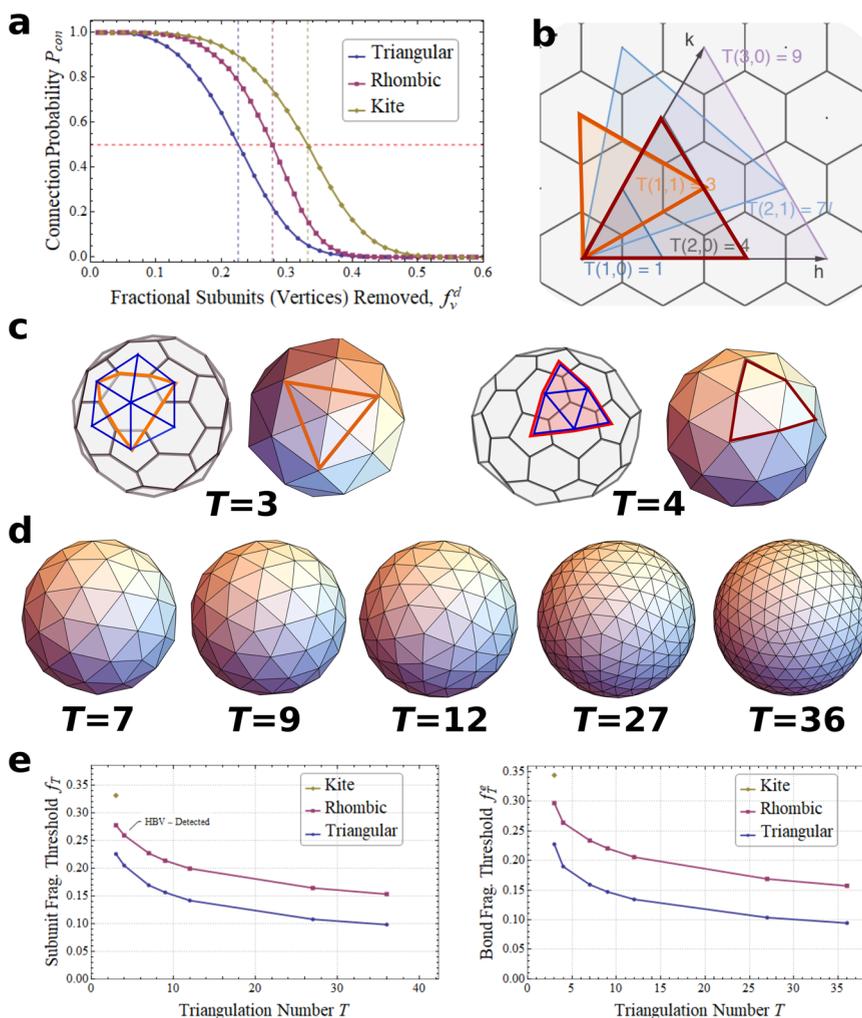
## RESULTS AND DISCUSSION

### Graphs as Topological Descriptors of Viral Capsids.

Viral capsid architectures are modeled in terms of surface tessellations termed tilings, in which tiles indicate capsid building blocks (capsomers). For the smallest icosahedral capsids formed from 60 copies of a single capsid building block, there is only one possible triangular layout that is fully

determined by icosahedral symmetry. The next larger capsids according to Caspar and Klug theory have triangulation number  $T = 3$  and are formed from  $60T = 180$  proteins. These include the architectures of plant viruses, such as Cowpea Chlorotic Mottle virus (CCMV), that are used in virus nanotechnology.<sup>20</sup> According to the classification of virus architecture based on the generalized quasi-equivalence principle,<sup>18</sup> there are three possible, topologically distinct layouts for capsids with this stoichiometry. Two of them correspond to tilings formed from 60 tiles, each representing three protein subunits: a triangular tile as seen in Pariacoto virus (1f8v, Figure 1A top), and a kite-shaped tile as in Tobacco Ringspot virus (1a6c, Figure 1B top). In addition, there is a capsid design formed from 90 rhombic tiles, each representing two protein subunits, as seen in bacteriophage MS2 (2ms2, Figure 1C top). These capsids are formed from one type of protein subunit, but these subunits can be in distinct conformations. Icosahedral symmetry constrains their positions in the capsid but allows for up to three distinct conformers in a  $T = 3$  architecture. As a consequence, the interfaces between tiles are not identical. Each triangular and kite tile accommodates proteins in three distinct conformations, and rhomb tilings are formed from two distinct types of protein dimer, a symmetric, and an asymmetric one, which are composed of protein subunits in different conformations.

The intersubunit bond network, *i.e.*, the topology of the network describing the contacts between adjacent subunits, corresponds to the dual tiling (Figure 1, bottom row), with vertices  $V_0$  (yellow) representing individual tiles (capsomers), and edges  $E_0$  interfaces between adjacent tiles. We use the graph  $\mathcal{G}$  corresponding to the dual tiling as a topological descriptor of the capsid. The vertices and edges form sets from which  $i$  elements may be randomly removed in order to probe the capsid's resilience to fragmentation using percolation theory. As in previous work, we refer to a complete capsid as  $\mathcal{G} = \mathcal{G}_0$  and any partially disassembled capsid randomly missing  $i$  tiles (vertices  $v$ ) as  $\mathcal{G}_i^v$ .<sup>19</sup> The fraction of deleted vertices is thus  $f_v^d = i/V_0$  and the fraction remaining is  $p_v^v = 1 - f_v^d$ , both of which are symmetric common variables in percolation theory. We use here the fraction deleted,  $f_v^d$ . Similarly, we refer to a capsid with  $j$  randomly broken bonds as  $\mathcal{G}_j^e$  and denote the fraction of broken



**Figure 2.** Some viral architectures are less stable and more prone to fragmentation. (a) The rapidly decaying, sigmoidal probability,  $P_{con}$ , that a given  $T = 3$  architecture is still connected, as a function of the fraction  $f_v^d$  of subunits removed, computed by averaging 100,000 Monte Carlo replicates per data point. The fragmentation threshold of each viral blueprint is indicated by gridlines. (b) The  $T$ -number construction for virus tilings according to Caspar and Klug.<sup>16</sup> An equilateral triangle, representing one of the 20 triangular faces of an icosahedron, is embedded into a hexagonal lattice such that the corners of the triangle are located in the centers of hexagons. The triangles corresponding to a  $T = 3$  and  $T = 4$  embedding are shown in orange and red, respectively. (c) The spherical architecture corresponding to the  $T = 3$  and  $T = 4$  embeddings are shown together with their corresponding dual triangulations to their right; the orange, and respectively red, triangles from (b) are shown superimposed. (d) Structures of the homogeneously tiled virus architectures, for which subunit and bond fragmentation thresholds ( $f_T$  and  $f_T^e$ ) have been computed (Table 1), demonstrating their dependence on capsid size. (e) The subunit (left) and bond (right) fragmentation thresholds for selected kite, rhombic, and triangular tilings of sizes up to  $T = 36$ . The subunit fragmentation threshold for rhombic  $T = 4$  tilings has already been experimentally confirmed for HBV;<sup>7,19</sup> others remain to be observed. The kite tiling is the most stable, followed by rhombic tilings, while triangular tilings are the most prone to fragmentation.

bonds (edges  $e$ ) as  $f_e^d = j/E_0$ . Note that the perforated capsid  $G_i^v$  ( $G_j^e$ ) can fragment into two or more separated clusters. These are then referred to as distinct “connected components” following terminology in graph theory.

**Percolation Theory for Viral Capsids.** Capsid fragmentation occurs *via* any dissociation of single units (tiles) or smaller clusters from the bulk of the capsid and is followed quickly by complete dissociation and breakdown of the remaining long-range connectivity. In order to profile this behavior, we have numerically determined the subunit and bond inverse fragmentation thresholds ( $f_T$ ) as an ensemble average over different stochastic simulations for each of the three geometrically and topologically distinct  $T = 3$  capsid architectures shown in Figure 1.

The predicted probability  $P_{con}(f_v^d)$  that a capsid missing a given fraction of subunits will remain connected shows

sigmoidal decay (Figure 2a), revealing a fairly abrupt transition from an intact, perforated capsid to dissociated fragments. The point at which the capsid is, on average, fragmented corresponds to the fragmentation threshold  $f_T$ , defined by  $P_{con}(f_v^d) = 0.5$ . Notably, the  $f_T$  of the triangular, kite, and rhombic  $T = 3$  architectures are distinct, with values of  $f_T = (0.226, 0.331, 0.278)$ , respectively, suggesting that the corresponding capsids exhibit different propensities for fragmentation. We converted the traditional percolation thresholds for the infinite planar lattices, from which our spherical lattices have been derived, into deletion fractions,  $p_c \approx (0.303, 0.378, 0.347)$ ,<sup>21–24</sup> to make them directly comparable to the fragmentation thresholds computed here. We note that the relative values are similar for the different lattice types, with the triangulation scoring lowest, and the kite tiling highest. Differences in absolute values are likely due to the fact that we are penalizing against the exclusion of singlets. Our

**Table 1. Fragmentation Thresholds  $f_T$  and  $f_T^c$  for Different Tiling Types and Sizes**

tiling type	$T$ -number	$f_T$	$f_T^c$
triangular	3	0.226	0.228
	4	0.205	0.19
	7	0.169	0.159
	9	0.156	0.147
	12	0.142	0.134
	27	0.108	0.104
	36	0.098	0.094
rhombic	3	0.278	0.297
	4	0.26	0.264
	7	0.227	0.234
	9	0.214	0.221
	12	0.199	0.206
kite	27	0.164	0.169
	36	0.153	0.157
	3	0.331	0.344

results are intuitive: the fragmentation threshold  $f_T$  (and percolation threshold) of a given lattice scales with the subunit valency or number of bonds per subunit, *i.e.*,  $f_T \propto k_0$ , where  $k_0 = 2\left(\frac{E_0}{V_0}\right)$ . Indeed, we observe successively higher thresholds in the tetravalent (4-connected) graphs associated with the rhombic and kite tilings, than in the graph of the trivalent (3-connected) triangular tiling.

The kite tiling is the most stable of the three homogeneously tiled capsid types with respect to random subunit removal and is the only one that is resistant to fragmentation beyond 40% subunit removal ( $f_v^d = 0.40$ ). This is mirrored by the bond fragmentation threshold based on random breakage of bonds rather than removal of tiles, which is  $f_T^c = (0.208, 0.318, 0.288)$  for the triangular, kite, and rhombic tiling, respectively. We again compare with the standard percolation thresholds,  $p_c^e = (0.347, 0.475, 0.476)$ .<sup>21–24</sup> We observe that in both cases the triangulation has the lowest value, whereas the two four-coordinated tilings, the kite and rhombic tiling, both have higher ones. We note that in the spherical tilings, there is a distinction between the fragmentation thresholds of these two four-coordinated tilings that is not visible in the percolation thresholds of the corresponding planar lattices. Interestingly, regarding initial fragmentation, the kite tiling remains the most stable tiling also with respect to random bond breakage.

We repeated the same construction for larger polyhedra that can be derived from the same planar lattices following the procedure introduced by Twarock and Luque.<sup>18</sup> For the hexagonal lattice, the corresponding polyhedra are known as Goldberg polyhedra or geodesic icosahedra. They are classified in terms of the triangulation number<sup>16</sup> ( $T$ -number)  $T$ , which specifies the position of one triangular face of an icosahedron in a hexagonal lattice embedding (Figure 2b). Each triangle is fully determined by  $T = h^2 + hk + k^2$ , with two integer numbers  $h$  and  $k$ . The latter represent steps between midpoints of adjacent hexagons along two vectors  $\vec{h}$  and  $\vec{k}$ , respectively, that intersect at the midpoint of a hexagon at an angle of  $\pi/3$ , cutting through two adjacent edges. The examples of  $T = 3$  and  $T = 4$  architectures are shown in Figure 2c. The Goldberg polyhedra have 12 pentagonal and  $10(T - 1)$  hexagonal faces.<sup>16</sup> Their duals, triangulations called deltahedra, are classified in terms of  $T$ , which indicates how many facets tile, by area, one of the 20 triangular faces of the icosahedral reference frame. These

polyhedra have  $20T$  triangular facets and  $30T$  edges (bonds). A similar approach based on the trihexagonal lattice results in rhombic tilings formed from  $30T$  rhombic faces and  $60T$  bonds.<sup>18</sup> The subunit and bond  $f_T$ s for homogeneously tiled capsids up to  $T = 36$  are provided in Table 2, demonstrating that this trend also persists for larger capsids. The tiling type, or capsomer geometry, is thus a determinant of robustness against disassembly.

**Table 2. Parameters Directly Relevant to Our Percolation Theory Model**

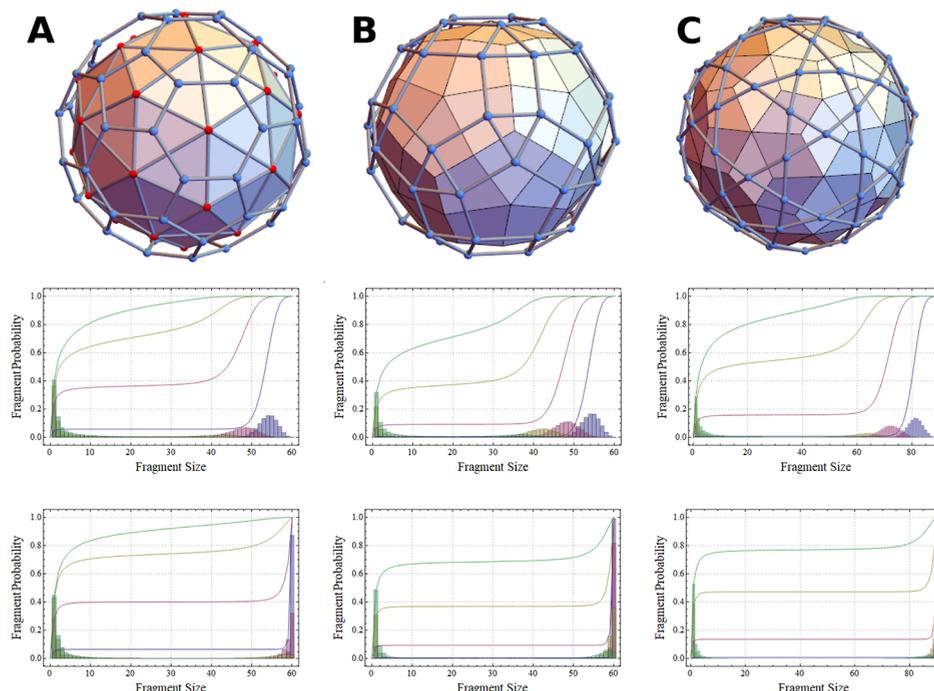
parameter	definition
$k_0$	initial capsid tiling's subunit valency (edges per vertex)
$f_v^d$	fraction of vertices/subunits deleted (independent variable)
$f_e^d$	fraction of edges/bonds deleted (independent variable)
$p_V$	fraction of vertices/tiles remaining
$p_E$	fraction of edges/bonds remaining
$P_{con}$	probability remaining subunits are connected (dependent variable)
$f_T$	vertex fragmentation threshold (where $P_{con}(f_v^d) = 0.5$ )
$f_T^c$	bond fragmentation threshold (where $P_{con}(f_e^d) = 0.5$ )
$p_c$	vertex percolation threshold
$p_c^e$	bond percolation threshold

**Implications for Capsid Stability.** In previous work, we introduced a closed-form empirical equation for general graphs, quantifying capsid stability in terms of the fraction  $f_v^d$  of randomly removed vertices. It characterizes the fraction of remaining bonds (edges)  $p_E$  as a function of the remaining fraction of subunits (or vertices)  $p_V = 1 - f_v^d$ :

$$p_E = p_V^2 = (1 - f_v^d)^2 \quad (1)$$

Note that subunit valency is not explicitly contained in this equation. Thus, differing fragmentation thresholds between tilings correspond to different fractions of preserved bonds at each threshold, implying that the impact of capsid geometry on stability extends beyond bond valency differences in the corresponding intersubunit bond networks. Indeed, while higher subunit valency enhances capsid stability and increases the fragmentation threshold, fragmentation appears to also be size-dependent. This can be seen from the following argument. The  $T = 3$  kite tiling has 60 tetravalent subunits and according to our analysis is more stable than the tetravalent 90-subunit rhombic tiling. However, for any given  $T$ -number, there are consistently 50% more subunits in the rhombic tilings, so that the ratio of building blocks alone cannot explain the shift in the fragmentation threshold seen in Figure 2e, where the values of the subunit (and bond) fragmentation thresholds are plotted against capsid size (in terms of the  $T$ -number). This implies that the stability difference between the rhombic and kite tilings is due to both the differing topology and the dependence of the subunit (and bond) fragmentation thresholds on capsid size.

The observed size dependence is, in part, due to an increase in the potential for exclusion of small clusters and single subunits from the bulk of the capsid: as the number of subunits increases, so does the probability of excluding small clusters. However, this size dependence appears to plateau at a value determined by the topology of the lattice. The triangular tiling, having lower subunit valency, is less resilient to fragmentation, particularly at higher  $T$ -numbers. By contrast, fragmentation of a rhombic tiling at  $T = 36$  requires removal of a  $\approx 50\%$  larger fraction of its subunits than for a triangular tiling, while it requires only a  $\approx 20\%$



**Figure 3.** Fragment size distributions for different capsid types: (a) triangular, (b) kite, and (c) rhombic tiling design. Middle and bottom row: The histogram probability density functions (PDFs) of fragments of various sizes with normally distributed  $f_v^d$  corresponding to 10% (blue), 20% (purple), 30% (gold), and 40% (green) subunit removal from the respective capsids/tilings (middle row) and edge removal from the corresponding interaction networks (bottom). These curves exclude the explicitly removed subunits and are overlaid with the cumulative distribution functions (CDFs), the slope of which indicates the fraction of species of that size. Interestingly, random bond breakage does not exhibit the same gradual decrease in fragment size until complete dissociation, as is observed experimentally for the dissociation of passivated subunits. Our theory predicts abrupt fragmentation beyond the fragmentation threshold and identifies expected concentrations that are experimentally measurable.

larger fraction to disassemble a  $T = 3$  rhombic tiling over a triangular one. Given the need for viral capsids to provide adequate protection for the genome between rounds of infection, this may explain the increasing relative abundance of rhombic tilings in nature as the size of the virus (in terms of its  $T$ -number) increases. While there are examples of triangular and rhombic tilings at  $T = 4$ , rhombic tilings dominate from  $T = 7$ .

**Fragment Size Distributions.** A capsid disassembly experiment, such as that implemented with HBV, is expected to result in a distribution of perforated viruses missing a mean number  $i$  (or fraction  $f_v^d$ ) of their subunits.<sup>7</sup> The precise numbers of distinct products are determined by experimental conditions such as denaturant concentration or the relative ratio of regular versus passivated subunits.<sup>7,19</sup> In regimes sufficiently below the fragmentation threshold  $f_T$ , this results in a binomial distribution of clusters centered around the mean number of removed subunits  $i$ . Before taking into account thermodynamic and kinetic effects that may skew this distribution, a greater topologically induced shift will occur due to significant fragmentation at, and just beyond,  $f_T$ . Upon fragmentation, the binomial cluster size distribution will break down rapidly as singlets and small clusters dissociate from the larger whole, followed by complete disassembly. This can be seen in Figure 3A–C, where the fragmentation behavior of the different tilings is characterized by assessing the expected fragment size distributions following random removal of a given fraction of either passivated subunits or individually broken bonds.

The probability density function (PDF) histograms and overlaid cumulative distribution functions (CDFs) are based on the binned results of typically 10 million partially dissociated

capsids per assessed distribution. They illustrate the empirical fragment size distributions  $P(N)$  at different degrees of fractional subunit and bond removal  $f_v^d$  (and  $f_e^d$ ) in the vicinity of the respective fragmentation thresholds. Note that in the case of subunit passivation and removal ( $f_v^d$ ; cf. Figure 3 (middle row)) passivated subunits are excluded from the distribution, whereas for bond breakage ( $f_e^d$ ; cf. Figure 3 (bottom row)) all subunits are accounted for. The graphs reveal the expected relative concentrations of fragments of various sizes upon random removal of subunits from the three distinct types of  $T = 3$  tilings in Figure 1. These four distributions are computed for successively higher subunit removal fractions,  $f_v^d = 10\%$  (blue), 20% (purple), 30% (gold), and 40% (green), and are shown roughly centered about  $f_T$ . Analogous results for random bond breakage are shown in Figure 3 (bottom row) based on the same removal values for  $f_e^d$ . Note that the only virus architecture with appreciable numbers of large clusters beyond 40% subunit removal is the kite tiling. This implies that this capsid design is much more stable than those conforming to other tiling types. This may account for the relatively rare occurrence of this tiling type in nature, and could perhaps be an indicator that this high degree of stability is not conducive to genome release.

Our predicted fragmentation threshold for HBV can be directly compared to experimental outcomes, as we have done previously in our study of the HBV capsid.<sup>7,19</sup> In these experiments, the disassembly of HBV capsids was studied in which 240 C-terminal truncated HBV monomers, with a tunable fraction passivated to disable covalent bond formation, were organized in dimeric (groups of two) subunits according to a predominantly  $T = 4$  surface rhombic tiling. Disassembly was

triggered by addition of a mild denaturant, interrupting the remaining comparably fragile hydrophobic contacts. The maximal fraction of subunits that could be removed before capsid fragmentation occurred was identified, and agreed well with the predicted value of the fragmentation threshold ( $f_T$ ) of approximately 26%.<sup>19</sup> This threshold was also observed for capsids assembled from passivated subunits only *via* the titration of mild amounts of denaturant. This strongly suggests that the theoretical predictions made here are robust against the specifics of the experimental setup.<sup>7</sup> They are thus of generic interest in applications to molecular breadboard and porosity-tuned nanotechnology using other viral capsids with different lattice blueprints.

## CONCLUSION

A virus' propensity for fragmentation is important for cargo release. The interplay of mutation and selection in viral evolution has resulted in capsids that balance stability, in order to provide sufficient protection for their genomic cargoes, and instability, in order to enable their timely release into the host cell environment. As we have shown here, this delicate balance hinges on capsid architecture and depends crucially on tiling type and the topology of the associated interaction network.

We demonstrate this explicitly here by computing subunit and bond fragmentation thresholds for the three  $T = 3$  capsid types that correspond to the smallest nontrivial Caspar and Klug capsid layouts in virology and include architectures that are currently exploited in virus nanotechnology. One of the unexpected conclusions from our work is that virus capsids organized according to a kite tiling are a standard deviation more stable, and triangular tilings a standard deviation less stable, than rhombic tilings. This may account for the shift from triangular to predominantly rhombic tiling architectures in larger viruses in nature, suggesting that capsid stability could be a driver for the evolution of specific protein subunit architectures.

Our model is based on a number of simplifying assumptions, in particular neglect of the thermodynamic and kinetic complexities of other models. Frameworks based on reaction kinetics and thermodynamics often use (large systems of) differential equations or employ more complex molecular dynamics simulations, respectively.<sup>25–27</sup> Differential equation models include subunit valency and combinatorics capturing one or a few static bond energy value(s), as well as reaction rate parameters, in order to model the kinetic approach to thermodynamic equilibrium. Molecular dynamics and Monte Carlo simulations, on the other hand, replace the static bond energy, as well as the implicit representation of subunit geometry, with explicit geometric subunit models and thermodynamic interaction potentials, as well as diffusive dynamics, in order to analyze the assembly kinetics and its long-term behavior. In contrast to these comparatively complex models, the percolation model presented here retains only the topological contributions from the subunit and bond network derived from geometric principles alone.<sup>16,18</sup> The simplicity of these comparably limited, strictly topological assumptions is reflected in the ease of the percolation model's implementation, which is much simpler than large-scale reaction kinetic models and molecular dynamics simulations. Nevertheless, this strictly topological percolation model and its prediction of the fragmentation threshold of HBV is in excellent agreement with experiment as outlined above,<sup>7,19</sup> demonstrating that our approach captures essential features of capsid fragmentation.

Indeed, our approach makes predictions that are testable experimentally and will enable experimentalists to detect the equivalent thresholds for any virus of interest.

We also note that the percolation threshold  $p_c$  studied in the broader percolation theory literature corresponds to the point of disruption of any long-range connectivity, that is, disappearance of any cluster of the order of system size, while only allowing the remaining smaller part of the system to be discontinuous. The more lenient percolation threshold  $f_T < p_c$  corresponds to long-range spanning percolation. This is comparable to the onset of fluid flow through a large-scale cluster spanning from one side of a 2D lattice medium to another, without assuming global connectivity of the lattice (here, of the remaining capsid). We argue that  $f_T$  is indeed the biologically and experimentally relevant quantity, as thermodynamic breakup is likely to occur much beyond the initial fragmentation threshold. This is consistent with our previous work on the fragmentation threshold of Hepatitis B virus, validated by single-particle detection methods with near single-subunit resolution, such as Charge Detection Mass Spectroscopy (CDMS) and nano-fluidics.<sup>7,19</sup> Such methods detect any dissociating singlets, much like the  $f_T$  metric itself. The observed sensitivity to the possibility of fragmented singlets or small clusters is also consistent with the size dependence of the fragmentation threshold  $f_T$  established here, which is not the case for the percolation threshold  $p_c$  in general.

The theory presented here provides a guide for applications in nanotechnology. We recently predicted and demonstrated that individual P22 VLP nanoreactors may be hierarchically assembled into ordered arrays *via* the use of small, oppositely charged linkers.<sup>8</sup> Such close-packed superlattices have also been generated for cowpea chlorotic mottle virus (CCMV),<sup>28–31</sup> which corresponds to one of the  $T = 3$  structures analyzed here. Close-packing is capable of increasing the local VLP concentration by several hundred fold, thus enhancing the catalytic activity of the cargo.<sup>2</sup> Enhanced catalysis is, in part, due to the porosity of the VLPs to diffusive small molecules. The tunable presence of holes in a VLP surface is likely to increase porosity, enabling the diffusion of small molecules, and thus allowing tuning of the catalytic activity of hierarchical VLP assemblies.<sup>2,8,32</sup> Such porosity tuning is already employed in the generation of porous coordination polymers for purposes of small molecule storage, separation, and catalysis.<sup>33</sup> Our results enable a better understanding of how (and if) the desired porosity may be achieved in biomimetic systems.

The resistance to fragmentation and disassembly of viruses also confers many other technological advantages to VLPs. The maximal number of subunits that may be removed, *i.e.*, the number of "holes" that can be "punched" into its surfaces before it fragments, informs the use of subunit removal and replacement strategies to tune a VLP's properties. For example, the ability to regulate the number of perforations (or modified subunits) constituting the VLPs will likely enable tuning of their elastic properties, which are commonly investigated using atomic force microscopy (AFM),<sup>34–38</sup> and play a vital role in virus assembly.<sup>11,12,39</sup> This type of subunit removal potentially enables particle shape to be controlled more readily by reducing the elastic moduli of the VLPs closer to regimes demonstrated to be susceptible to deformation.<sup>34,35,40,41</sup> Functionalization may be possible, as the reversibility of assembly and disassembly has enabled the refilling of up to this maximum number of missing subunits with different, potentially functionalizable subunits in so-called chimeric molecular breadboards, such as those based

on the HBV nucleocapsid<sup>7</sup> or the P22 bacteriophage capsid.<sup>13</sup> Indeed, other forms of surface functionalization of fully assembled capsids, and even simple solution additives, have already been shown to modulate the VLP elastic response.<sup>37,38</sup> Array formation has also been shown to be sensitive to the net charge of the virus, which could change if mutated subunits with fewer charge moieties were substituted into the chimera.<sup>8,32</sup>

The extent to which missing subunits and VLP breadboard (subunit replacement) methods will be successful in tuning porosity, elasticity, charge, and other surface properties is determined, in part, by the maximal number of subunits that may be removed without inducing collapse of the entire VLP. Our results therefore not only identify drivers of viral evolution, favoring certain capsid architectures in nature, but also provide a guide for the exploitation of VLPs in virus nanotechnology, enabling better control of the biological properties of VLP-based biomimetic materials.

## METHODS

**Mathematical Representation of Viral Capsid Architecture as Tilings.** Blueprints of icosahedral viral capsids abide to an overarching design principle based on Archimedean lattices.<sup>18</sup> We focus here on those capsid geometries in this classification that can be constructed from a single subunit type, which are the (6, 6, 6), (3, 6, 3, 6), and (3, 4, 6, 4) lattices. These geometric models refine the Caspar and Klug classification of virus architecture, in which capsids are described in terms of the triangulation number  $T$ , where  $T = h^2 + hk + k^2$  for  $h$  and  $k$  non-negative integers. In their theory, capsids are formed from  $60T$  proteins that are organized into 12 clusters of five and  $10(T - 1)$  clusters of six proteins. There is only one geometric blueprint for a  $T = 3$  capsid formed from 180 proteins. However, Archimedean lattice theory identifies the three distinct types of models in Figure 1 with different properties in terms of tile numbers and interaction networks.

**The Fragmentation Threshold.** Subunit and bond fragmentation thresholds were computed numerically for each tiling type. After sufficient (typically 100 thousand) replicates we observed a smooth, sigmoidal decrease in the probability that all remaining subunits are connected,  $P_{con}$ , indicative of a phase transition from a connected “holey” capsid to individual fragments. We define the  $f_T$  as the critical percentage of subunits (or bonds) deleted at which the system is, on average, disconnected ( $P_{con} = 0.5$ ). The code is available from GitHub: <https://github.com/MathematicalComputationalVirology/VirusPercolationTheory.git>

**Overview of Model Parameters.** The parameters in Table 2 are directly relevant to our percolation theory model. The model has not otherwise been parametrized prior to numerical estimates of the final quantities of interest upon random vertex/edge deletion.

## AUTHOR INFORMATION

### Corresponding Authors

Nicholas E. Brunk – Wolfram Research, Champaign, Illinois 61820, United States; VeriSIM Life, San Francisco, California 94104, United States; Intelligent Systems Engineering, Indiana University, Bloomington, Indiana 47408, United States; [orcid.org/0000-0001-5444-799X](https://orcid.org/0000-0001-5444-799X); Email: [nbrunk@wolfram.com](mailto:nbrunk@wolfram.com)

Reidun Twarock – Departments of Mathematics and Biology, York Cross-disciplinary Centre for Systems Analysis, University of York, York YO10 5GE, U.K.; [orcid.org/0000-0002-1824-2003](https://orcid.org/0000-0002-1824-2003); Email: [reidun.twarock@york.ac.uk](mailto:reidun.twarock@york.ac.uk)

Complete contact information is available at: <https://pubs.acs.org/10.1021/acsnano.1c01882>

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

R.T. acknowledges funding via an EPSRC Established Career Fellowship (EP/R023204/1), a Royal Society Wolfson Fellowship (RSWF/R1/180009), and a Joint Wellcome Trust Investigator Award with Peter Stockley from the University of Leeds (110145 & 110146). N.E.B. acknowledges funding via Wolfram Research.

## REFERENCES

- (1) Czapar, A. E.; Steinmetz, N. F. Plant Viruses and Bacteriophages for Drug Delivery in Medicine and Biotechnology. *Curr. Opin. Chem. Biol.* **2017**, *38*, 108–116. Next Generation Therapeutics
- (2) Uchida, M.; McCoy, K.; Fukuto, M.; Yang, L.; Yoshimura, H.; Miettinen, H. M.; LaFrance, B.; Patterson, D. P.; Schwarz, B.; Karty, J. A.; Prevelige, P. E.; Lee, B.; Douglas, T. Modular Self-Assembly of Protein Cage Lattices for Multistep Catalysis. *ACS Nano* **2018**, *12*, 942–953.
- (3) Douglas, T.; Young, M. Host–Guest Encapsulation of Materials by Assembled Virus Protein Cages. *Nature* **1998**, *393*, 152–155.
- (4) Douglas, T.; Young, M. Viruses: Making Friends with Old Foes. *Science* **2006**, *312*, 873–875.
- (5) Glotzer, S. C. Assembly Engineering: Materials Design for the 21st Century (2013 P.V. Danckwerts lecture). *Chem. Eng. Sci.* **2015**, *121*, 3–9. 2013 Danckwerts Special Issue on Molecular Modelling in Chemical Engineering
- (6) Mateu, M. G. Assembly, Engineering and Applications of Virus-Based Protein Nanoparticles. In *Protein-Based Engineered Nanostructures*; Cortajarena, A. L., Grove, T. Z., Eds.; Advances in Experimental Medicine and Biology 940; Springer: Springer International Publishing Switzerland, 2016; pp 83–120.
- (7) Lee, L. S.; Brunk, N. E.; Haywood, D. G.; Keifer, D.; Pierson, E.; Kondylis, P.; Wang, J. C.-Y.; Jacobson, S. C.; Jarrold, M. F.; Zlotnick, A. A Molecular Breadboard: Removal and Replacement of Subunits in a Hepatitis B Virus Capsid. *Protein Sci.* **2017**, *26*, 2170–2180.
- (8) Brunk, N. E.; Uchida, M.; Lee, B.; Fukuto, M.; Yang, L.; Douglas, T.; Jadhao, V. Linker-Mediated Assembly of Virus-Like Particles into Ordered Arrays via Electrostatic Control. *ACS Applied Bio Materials* **2019**, *2*, 2192–2201.
- (9) Kegel, W. K.; van der Schoot, P. Competing Hydrophobic and Screened-Coulomb Interactions in Hepatitis B Virus Capsid Assembly. *Biophys. J.* **2004**, *86*, 3905–3913.
- (10) van der Schoot, P.; Bruinsma, R. Electrostatics and the Assembly of an RNA Virus. *Phys. Rev. E* **2005**, *71*, 061928.
- (11) Zandi, R.; van der Schoot, P.; Reguera, D.; Kegel, W.; Reiss, H. Classical Nucleation Theory of Virus Capsids. *Biophys. J.* **2006**, *90*, 1939–1948.
- (12) Erdemci-Tandogan, G.; Wagner, J.; van der Schoot, P.; Podgornik, R.; Zandi, R. Effects of RNA Branching on the Electrostatic Stabilization of Viruses. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2016**, *94*, 022408.
- (13) Kang, S.; Uchida, M.; O’Neil, A.; Li, R.; Prevelige, P. E.; Douglas, T. Implementation of P22 Viral Capsids as Nanoplatfoms. *Biomacromolecules* **2010**, *11*, 2804–2809.
- (14) Schrenk, K.; Hilário, M.; Sidoravicius, V.; Araújo, N.; Herrmann, H.; Thielmann, M.; Teixeira, A. Critical Fragmentation Properties of Random Drilling: How Many Holes Need to be Drilled to Collapse a Wooden Cube? *Phys. Rev. Lett.* **2016**, *116*, 055701.
- (15) Crick, F. H.; Watson, J. D. Structure of Small Viruses. *Nature* **1956**, *177*, 473–5.
- (16) Caspar, D. L. D.; Klug, A. Physical Principles in the Construction of Regular Viruses. *Cold Spring Harbor Symp. Quant. Biol.* **1962**, *27*, 1–24.
- (17) Twarock, R. A. Tiling Approach to Virus Capsid Assembly Explaining a Structural Puzzle in Virology. *J. Theor. Biol.* **2004**, *226*, 477–482.
- (18) Twarock, R.; Luque, A. Structural Puzzles in Virology Solved with an Overarching Icosahedral Design Principle. *Nat. Commun.* **2019**, *10*, 4414.

- (19) Brunk, N. E.; Lee, L. S.; Glazier, J. A.; Butske, W.; Zlotnick, A. Molecular Jenga: The Percolation Phase Transition (Collapse) in Virus Capsids. *Phys. Biol.* **2018**, *15*, 056005.
- (20) Koudelka, K. J.; Pitek, A. S.; Manchester, M.; Steinmetz, N. F. Virus-Based Nanoparticles as Versatile Nanomachines. *Annu. Rev. Virol.* **2015**, *2*, 379–401.
- (21) Suding, P. N.; Ziff, R. M. Site Percolation Thresholds for Archimedean Lattices. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1999**, *60*, 275.
- (22) Parviainen, R. Estimation of Bond Percolation Thresholds on the Archimedean Lattices. *J. Phys. A: Math. Theor.* **2007**, *40*, 9253–9258.
- (23) Sykes, M. F.; Essam, J. W. Exact Critical Percolation Probabilities for Site and Bond Problems in Two Dimensions. *J. Math. Phys.* **1964**, *5*, 1117–1127.
- (24) Ziff, R. M.; Suding, P. N. Determination of the Bond Percolation Threshold for the Kagomé Lattice. *J. Phys. A: Math. Gen.* **1997**, *30*, 5351.
- (25) Zlotnick, A. To Build a Virus Capsid: An Equilibrium Model of the Self Assembly of Polyhedral Protein Complexes. *J. Mol. Biol.* **1994**, *241*, 59–67.
- (26) Keef, T.; Micheletti, C.; Twarock, R. Master Equation Approach to the Assembly of Viral Capsids. *J. Theor. Biol.* **2006**, *242*, 713–721.
- (27) Perlmutter, J. D.; Perkett, M. R.; Hagan, M. F. Pathways for Virus Assembly around Nucleic Acids. *J. Mol. Biol.* **2014**, *426*, 3148–3165.
- (28) Kostiaainen, M. A.; Hiekkataipale, P.; Jose, A.; Nolte, R. J.; Cornelissen, J. J. Electrostatic Self-Assembly of Virus–Polymer Complexes. *J. Mater. Chem.* **2011**, *21*, 2112–2117.
- (29) Kostiaainen, M. A.; Hiekkataipale, P.; Laiho, A.; Lemieux, V.; Seitonen, J.; Ruokolainen, J.; Ceci, P. Electrostatic Assembly of Binary Nanoparticle Superlattices Using Protein Cages. *Nat. Nanotechnol.* **2013**, *8*, 52.
- (30) Liljeström, V.; Mikkilä, J.; Kostiaainen, M. A. Self-Assembly and Modular Functionalization of Three-Dimensional Crystals from Oppositely Charged Proteins. *Nat. Commun.* **2014**, *5*, 4445.
- (31) Kostiaainen, M. A.; Kasyutich, O.; Cornelissen, J. J.; Nolte, R. J. Self-Assembly and Optically Triggered Disassembly of Hierarchical Dendron–Virus Complexes. *Nat. Chem.* **2010**, *2*, 394.
- (32) Brunk, N. E.; Kadupitiya, J.; Uchida, M.; Douglas, T.; Jadhao, V. *Nanoparticle Assembly Lab*, 2019; <https://nanohub.org/resources/npassemblylab>; Version 1.2.1; 18 Nov 2020.
- (33) Spokoyny, A. M.; Farha, O. K.; Mulfort, K. L.; Hupp, J. T.; Mirkin, C. A. Porosity Tuning of Carborane-Based Metal–Organic Frameworks (MOFs) via Coordination Chemistry and Ligand Design. *Inorg. Chim. Acta* **2010**, *364*, 266–271.
- (34) Lidmar, J.; Mirny, L.; Nelson, D. R. Virus Shapes and Buckling Transitions in Spherical Shells. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2003**, *68*, 051910.
- (35) Šiber, A. Buckling Transition in Icosahedral Shells Subjected to Volume Conservation Constraint and Pressure: Relations to Virus Maturation. *Phys. Rev. E* **2006**, *73*, 061915.
- (36) Roos, W.; Bruinsma, R.; Wuite, G. Physical Virology. *Nat. Phys.* **2010**, *6*, 733.
- (37) Snijder, J.; Radtke, K.; Anderson, F.; Scholtes, L.; Corradini, E.; Baines, J.; Heck, A. J.; Wuite, G. J.; Sodeik, B.; Roos, W. H. Vertex-Specific Proteins pUL17 and pUL25 Mechanically Reinforce Herpes Simplex Virus Capsids. *J. Virol.* **2017**, *91*, e00123–17.
- (38) van Rosmalen, M. G.; Li, C.; Zlotnick, A.; Wuite, G. J.; Roos, W. H. Effect of dsDNA on the Assembly Pathway and Mechanical Strength of sv40 vp1 Virus-Like Particles. *Biophys. J.* **2018**, *115*, 1656–1665.
- (39) Li, S.; Roy, P.; Travesset, A.; Zandi, R. Why Large Icosahedral Viruses Need Scaffolding Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 10971–10976.
- (40) Jadhao, V.; Thomas, C. K.; de la Cruz, M. O. Electrostatics-Driven Shape Transitions in Soft Shells. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 12673–12678.
- (41) Brunk, N. E.; Jadhao, V. Computational Studies of Shape Control of Charged Deformable Nanocontainers. *J. Mater. Chem. B* **2019**, *7*, 6370–6382.