Tilable nature of virus capsids and the role of topological constraints in natural capsid design

Ranjan V. Mannige*
Department of Molecular Biology and Center for Theoretical Biological Physics, The Scripps Research Institute, 10550 North Torrey Pines Road, TPC 6, La Jolla, California 92037, USA

Charles L. Brooks III†
Department of Molecular Biology and Center for Theoretical Biological Physics, The Scripps Research Institute, 10550 North Torrey Pines Road, TPC 6, La Jolla, California 92037, USA
and Department of Chemistry and Biophysics Program, University of Michigan, 930 North University Avenue, Ann Arbor, Michigan 48104, USA

(Received 1 December 2007; published 1 May 2008)

Virus capsids are highly specific assemblies that are formed from a large number of often chemically identical capsid subunits. In the present paper we ask to what extent these structures can be viewed as mathematically tilable objects using a single two-dimensional tile. We find that spherical viruses from a large number of families—eight out of the twelve studied—qualitatively possess properties that allow their representation as two-dimensional *monohedral tilings* of a bound surface, where each tile represents a subunit. This we did by characterizing the extent to which individual spherical capsids display subunit-subunit (1) holes, (2) overlaps, and (3) gross structural variability. All capsids with T numbers greater than 1 from the Protein Data Bank, with homogeneous protein composition, were used in the study. These monohedral tilings, called *canonical capsids* due to their platonic (mathematical) form, offer a mathematical segue into the structural and dynamical understanding of not one, but a large number of virus capsids. From our data, it appears as though one may only break the long-standing rules of quasiequivalence by the introduction of subunit-subunit structural variability, holes, and gross overlaps into the shell. To explore the utility of canonical capsids in understanding structural aspects of such assemblies, we used graph theory and discrete geometry to enumerate the types of shapes that the tiles (and hence the subunits) must possess. We show that topology restricts the shape of the face to a limited number of five-sided prototiles, one of which is the "bisected trapezoid" that is a platonic representation of the most ubiquitous capsid subunit shape seen in nature (the trapezoidal jelly-roll motif). This motif is found in a majority of seemingly unrelated virus families that share little to no host, size, or amino acid sequence similarity. This suggests that topological constraints may exhibit dominant roles in the natural design of biological assemblies, while having little effect on amino acid sequence similarity.

DOI: 10.1103/PhysRevE.77.051902 PACS number(s): 87.10.−e

I. INTRODUCTION

The genomic material of a majority of viruses is enclosed and protected by spherical capsids. These spherical capsids are icosahedral in symmetry and are composed of protein subunits that are often chemically identical. The number of subunits that exist in a capsid (i.e., the size) is generally described by the triangulation (T) number, where a spherical capsid of triangulation number T will possess $60T$ subunits. The T number must satisfy the equation $T=h^2+hk+k^2$, where h and k are non-negative integers [1]. Often, capsid subunits of a specific virus family assemble into capsids of highly specific size or T number. For example, sesbania mosaic virus capsid subunits are known to spontaneously form $T=3$ spherical capsids [2], while the subunits belonging to the birnaviridae family form larger $T=13$ capsids [3]; different capsid sizes are formed on account of subunits interacting within the capsid in different but quasiequivalent manners. These rules of quasiequivalence, which are at the core of the description of the organization of such capsids, have been long standing and were proposed by Caspar and Klug in 1962 [1].

As indicated by the rules of quasiequivalence, even the smallest virus capsids must be made up of at least 60 protein subunits, each containing thousands of atoms themselves, making the analysis of structural and dynamical properties of these systems in all-atom form computationally difficult. For example, state of the art numerical simulations of a $T=1$ all-atom capsid were performed for more than 50 ns at the approximate rate of 1.1 ns per day on 48 processors [4]. Although a testament to parallel processivity, the simulation was also a testament to the current inadequacies, for virus capsids must be simulated for more than 1 ms (100 000 times more) to record important aspects such as assembly and structural changes. Coarse-graining of such all-atom systems (where atom clusters are grouped and treated as simpler pseudoatoms) reduces the system size drastically [5] but still falls short at the microsecond time scale (also, these systems pose the new problem of correct parametrization of the coarse-grain force field) [5]. Furthermore, biochemical and biophysical laboratory studies of important events such as capsid maturation and assembly offer few clues to the dynamic mechanisms involved. These complications have motivated a plethora of theoretical attempts aimed at understanding virus capsids using necessary and simplifying

*ranjan@umich.edu
†brookscl@umich.edu

TABLE I. The list of families studied along with their triangulation (T) numbers, average overlap_M percentage (% O), average percentage holes (% H), and average subunit variance in angstroms (V) within each family. The line divides the families into two groups: (1) families whose capsids may be representable by monohedral tilings (with relatively low % O, % H, and V), and (2) capsids that cannot be represented as monohedral tilings, i.e., capsids that possess holes, gross overlaps, and subunit variability.

Group	Family name	T	% O	% H	V
1	Tombusviridae	3	0.421	0.717	0.426
	Sobemoviridae	3	0.757	0.218	0.605
	Birnaviridae	13	0.350	0.870	0.648
	Nodaviridae	3	4.632	0.382	0.421
	Tymoviridae	3	0.242	1.455	1.075
	Siphoviridae	7l	3.860	2.777	1.138
	Bromoviridae	3	2.253	3.916	1.175
	Caliciviridae	3	7.002	2.502	1.189
2	Tetraviridae	4	12.504	0.001	1.052
	Hepadnaviridae	4	5.653	6.769	0.780
	Leviviridae	3	15.191	4.320	2.028
	Polyomaviridae	7d	16.943	5.033	2.524

geometric assumptions regarding the nature of capsid assemblies.

The use of simplified geometric models such as disks on a sphere [6,7], simple van der Waals spheres [8], Stockmayer fluids [9], trapezoidal subunits [10,11], tiles [12], and simple bonding units [13–15] has enabled the application of physics and mathematics in the exploration of various capsid phenomena such as assembly kinetics, capsid subunit stoichiometry, quasiequivalence, and assembly nucleation events. Although many of these models accurately describe specific *phenomenological* aspects of a capsid (such as disks on a sphere explaining the emergence of icosahedral symmetry [6,7]), a simplified capsid model that represents a broad array of properties of many natural capsids in an accurate manner (i.e., a model possessing "transferability") has yet to be described.

We are interested in understanding the extent to which simplified models may represent spherical capsids in nature. This problem will be addressed in the first half of the paper, where we will show that a large number of virus capsids ("group 1" capsids; see Table I) found in nature can, indeed, be modeled as simplified two-dimensional tilings of a bound surface by a single prototile, i.e., we will show that a large number of virus capsids may be represented by *monohedral tilings*. Capsids that follow these monohedral tilability rules (along with the quasiequivalence rules) we call *canonical capsids*. Mathematical studies of these monohedral tilings (canonical capsids), using tools such as tiling theory previously used in more specialized studies [12,16], are expected to relate the physical and structural properties obtained directly to a large number of capsids found in nature (corresponding to the group 1 capsids).

As an example of one mathematical application, the second part of our paper asks: What are the shapes available to a spherical capsid subunit as dictated by topological rules? Since the subunit is directly represented by a canonical capsid prototile, one only needs to ask how topology dictates the types of tiles that the canonical capsid may have. This is answered using graph theory, tiling theory, and the rules of spherical capsid equivalence laid down by Caspar and Klug [1]. From this, we show that the expected canonical capsid subunit shape strongly reflects the trapezoidal subunit shapes found in natural capsids.

II. MATERIALS AND METHODS

A. Virus capsids analyzed

For our analyses, we used all capsids present in the VIPERdb virus capsid repository (as of April 2007) [17] comprised of chemically identical subunits with triangulation numbers greater than 1. Capsids denoted by the following Protein Data Bank (PDB) IDs (also VIPERdb IDs) were used in the analysis (65 in number): 1aq3, 1aq4, 1auy, 1bms, 1c8n, 1cwp, 1ddl, 1dwn, 1dzs, 1e57, 1e7x, 1f15, 1f2n, 1f8v, 1frs, 1fr5, 1gav, 1gkv, 1gkw, 1ihm, 1js9, 1kuo, 1laj, 1mst, 1mva, 1mvb, 1ng0, 1nov, 1ohf, 1ohg, 1opo, 1qbe, 1qgt, 1qjz, 1sva, 1sid, 1sie, 1smv, 1u1y, 1w39, 1wce, 1x35, 1za7, 1zdh, 1zdi, 1zdj, 1zdk, 1zse, 2b2d, 2b2e, 2b2g, 2bbv, 2bu1, 2frp, 2fs3, 2fsy, 2ft1, 2gh8, 2ms2, 2tbv, 4sbv, 5msf, 6msf, 7msf, fhv. (The structure named fhv is not present in the PDB and was deposited into the VIPERdb as a personal communication.)

B. Definition of a monohedral tiling

We may say that a two-dimensional monohedral bound tiling is one where identically shaped (congruent) tiles come together onto a bound (topologically spherical) surface such that no tile-tile overlaps and holes are found. In slightly more specific terms, the term "monohedral two-dimensional tiling" refers to the classical *strongly balanced tiling* by a single prototile. Strongly balanced tiling means that each tile must

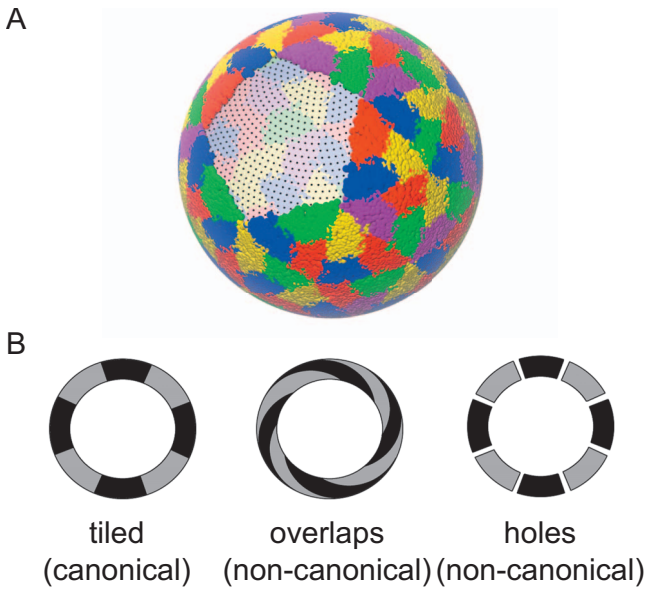

FIG. 1. (Color) Analysis of the tilability of a capsid. (a) The three-dimensional all-atom model is projected onto a sphere of average shell radius (colored to represent individual subunits) upon which a net, or dot matrix, is cast (visible in the subunit cleared area). These dots are used in calculating the (areawise) percentage of holes and overlaps in the capsid shell. (This projected capsid was derived from PDB ID 1smv). (b) The presence of either holes (right) or subunit-subunit overlaps (middle) will result in the inability to represent these structures as well-behaved two-dimensional tilings (left).

be a topological disk (polygon), the assembled tilings must represent a two-manifold (tiles must not overlap), edges cannot be disconnected (i.e., it is an edge-edge tiling), and they must be uniformly bounded and balanced (introduced by Grünbaum and Shepherd to preclude "paradoxical" tilings in their authoritative treatise on tilings and patterns [18]). Monohedrality imposes the need for just one tile shape to exist within the tiling; however, each instance of this shape need not be related by any symmetry operation. For an understanding of edge-to-edge monohedral tilings please refer to the review by Grünbaum and Shephard [19].

C. Characterizing tilability

We first need to characterize the extent to which virus capsids display holes and overlaps. Although this may be qualitatively done visually, we chose to develop a simple metric for quantitative characterization. In this method, we projected each protein atom present in the all-atom capsid structure onto a sphere whose radius equals the average radius of the capsid shell (depicted in Fig. 1). We then cast a ~ 1 Å^2 square-latticed net of dots—or a "dot matrix"—onto the shell and calculated the percentage of dots that were present within holes and subunit-subunit overlaps. The matrix created possessed no icosahedral symmetry and was not created by triangulations of an icosahedron (a heuristic approach was used instead). The volume of each subunit was defined by the van der Waals radii of the constituent atoms along with a 1.4 Å addition accounting for water. These percentages are obtained for each virus capsid. The density of the dots is large enough for a high-resolution characterization of holes and overlaps, especially given that the smallest atom in our structures is the carbon atom with radius ~ 1.7 Å.

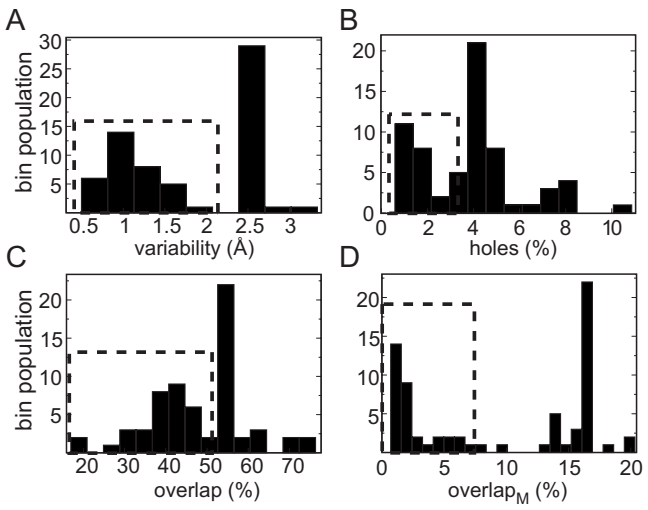

FIG. 2. There are a large number of spherical capsids, highlighted by the dashed rectangles, that possess one of the three requirements of monohedral tilability.

D. Characterizing monohedral tilability

Assuming that the capsid is tilable, monohedral tilability is found when the shape of each tile (or subunit) is the same (or congruent). To investigate tile congruence, we look at variability within subunits in a capsid, i.e., we structurally compare subunits within the asymmetric unit of the crystal structure to each other (since the asymmetric unit possesses the maximally different structures within the crystal structure). Note that our interest lies in characterizing structural changes in the *entire subunit*, and not localized conformational changes which alter the intersubunit interactions in an otherwise structurally rigid subunit, e.g., the order-disorder transitions in the tomato bushy stunt virus capsid (reviewed in [20]). This is because those structural changes may be manifested in the tiling as subunit-subunit dihedral angle changes.

In the final section of this paper, we use topology and tiling theory to compile a list of projected (two-dimensional) shapes available to the spherical virus capsid subunit.

III. RESULTS AND DISCUSSION

A. Tilability of natural spherical capsids

The extent to which capsids possess any of the three properties—(1) holes, (2) overlaps once projected onto a sphere, and (3) architectural variability—is used as a metric for tilability.

Using methods described in Sec. II, four values were calculated for each capsid and plotted as histograms in Fig. 2. The four graphs probe the following four properties: (a) monohedrality (subunit variability within a capsid measured by an averaged root mean squared deviation value in ang-

stroms), (b) the amount of breaks within the capsid shell (% holes), (c) the percentage of subunit-subunit overlap within the capsid, and (d) the percentage of *gross* subunit overlap, denoted by overlap_M (which was calculated by first shrinking each pruned subunit[1] by a scaling factor of 0.83 and then calculating the percentage overlaps without the 1.4 Å addition). The last graph was used to differentiate between those capsids that have normal overlaps—caused by interdigitation of neighboring amino acid residues into each other at the subunit-subunit interfaces—and gross structural (subunit-subunit) overlaps. It is clear that even in the best-behaved two-dimensional (2D) representable virus capsid, residue-residue interdigitations are inevitable; it is only the gross subunit-subunit overlaps that pose a hindrance to 2D tiling representations. In all of the metrics used, low values indicate that the capsid possesses little structural variability, negligible holes, or negligible overlaps.

It is immediately evident from the histograms that there is one group of capsids (highlighted by the dashed squares) where at least one of the three properties of monohedral tilability is possessed. The next question is whether some capsids possess all three properties (making them representable by monohedral tilings).

The average percent overlap_M, percent holes, and subunit structural variability were calculated for each virus family and these values were plotted against each other, resulting in the three graphs in Fig. 3(a)–3(c). These graphs indicate that all three properties are positively correlated with each other. Conversely, as we move away from any one of the three properties, the other two properties tend to weaken too, with the exceptions being the families tetraviridae and hepadnaviridae.

It is evident that there are capsids (belonging to families in group 2 in Table I) that present either high structural variability, subunit-subunit overlaps (tetraviridae), holes (hepadnaviridae), or all three characteristics (leviviridae and polyomaviridae) that prevent them from being represented as monohedral 2D bound tilings, or canonical capsids. A levivirus has been represented in Fig. 3(d) as an example of the holes and overlaps found in group 2 capsids.

Importantly, there are a large number of capsids (capsids belonging to 8 of the 12 families studied; the first group in Table I) that possess all three qualities of monohedral tilability and "reside" within all the dashed-line boxes in Fig. 2. These capsids may be represented by bound monohedral tilings that we call canonical capsids.

Although the bimodality of the histogram distributions in Fig. 2 (into group 1 and group 2) is evident, there is some group 1–group 2 overlap indicating that capsids close to the border (such as capsids belonging to bromoviridae and caliciviridae families of the group 1) may display subtle characteristics of the other class. This is expected when attempting to classify a biological system.

Our interest now is to show that predictions made on the *platonic* group 1 virus capsid—discrete mathematical models

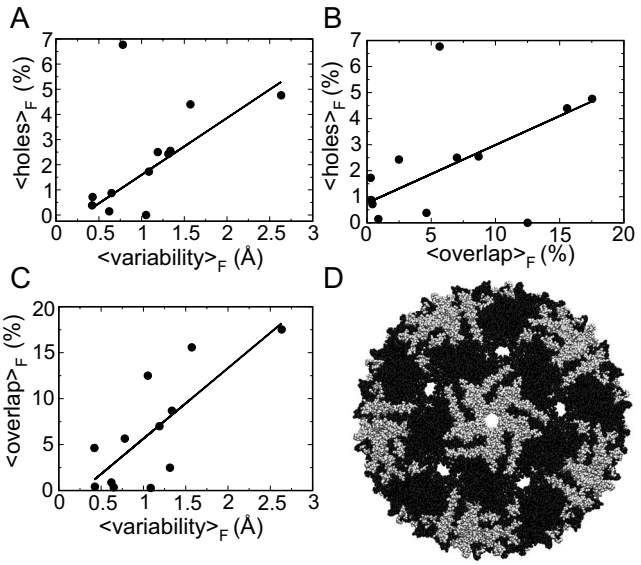

FIG. 3. (a)–(c) Dependence of one of three requirements for monohedral tilability plotted versus each of the others using family-averaged values. The solid lines are added to emphasize the trends. (d) An example of a capsid belonging to the Leviviridae family (PDB ID 1mst), which violates all of the requirements of monohedral tilability, as indicated by excessive overlap of black and white subunits and large holes at five- and pseudosixfold symmetry axes. (Subunit coloring: *A*,*B*, black; *C*, white.)

or canonical capsids—can, indeed, be related back to the capsids belonging to families in group 1 of Table I.

B. Definition of the canonical capsid

Canonical capsids are monohedral tilings or polyhedra where the tiles or faces within one polyhedron are related to each other by the rules of quasiequivalence [1]. Spherical capsids adhering to this paradigm possess $60T$ subunits, and may also be thought of as groupings of 12 pentamers and $10(T-1)$ hexamers. The manner in which one may go about making a Caspar-Klug capsid is by starting off with a hexagonal lattice and, based on a set of rules, converting specific hexagons into pentagons in a process that introduces curvature (or a disclination) to the surface. Please refer to [21] for a broader understanding of the geometric and physical basis of such transformations. With these concepts in mind, we can characterize the types of two-dimensional (projected) shapes that a subunit must possess in order to assemble into a canonical capsid of a specific T number.

C. Characterization of the subunit shape

A characterization of the types of shapes available to the canonical capsid subunit will be attempted by sequentially answering two more tractable questions. First, we will find the number of *edges* (σ) that the canonical capsid prototile may possess. Then we search for specific *tilings* (and their allowable shapes) that are in accordance with the canonical capsid definition and the allowable σ's. However, we need to formulate a representation of the subunit that enables these questions to be answered.

[1]"Pruned," here, means that the those amino acids within a subunit that undergo order-disorder transitions [20], which partake in the modifying of one interface, are ignored in the assay.

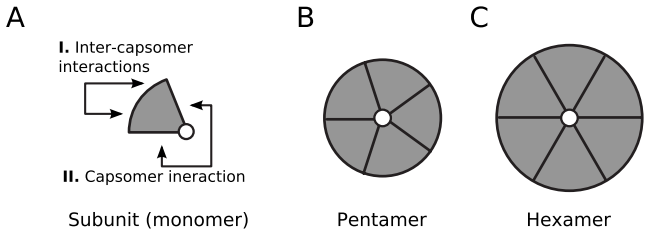

FIG. 4. Abstract representation of a canonical capsid subunit, represented here in subunit form along with the types of subunit-subunit interactions (a) and in capsomere (hexamer and pentamer) arrangements that are formed from type II interactions (b),(c). The final capsid is composed of 12 pentamers and $10(T-1)$ hexamers that interact with each other via type I interactions [1]. For our studies, in order to not constrain the number of subunit vertices (and hence edges), we define the curved edge as one with a potentially unlimited number of vertices (with a lower bound of 2).

Figure 4(a) shows a representation of the prototile with one quasiequivalent vertex (circled), two straight edges, and one curved edge. The curved edge may possess any number of trivalent vertices (allowing this structure to have an arbitrary number of edges and vertices). This has been done so as not to limit the number of shapes of the tile (dictated by the number of vertices or edges). Being the center of either a pentamer or hexamer [Figs. 4(b) and 4(c)], each quasiequivalent vertex may be either five- or six-valent. The number of vertices and edges (σ) may be dependent on various properties of the capsid such as the size or T number.

The ability to describe capsids as monohedral tilings (or polyhedra) and our description of a σ-unrestricted subunit allow the canonical capsid to be systematically analyzed using simple topological tools such as Euler's polyhedral formula.

1. The number of sides (σ)

All canonical capsids can be thought of, graph theoretically, as triangulations of the icosahedron. As the icosahedron is convex, it can be projected as a planar graph onto a two-dimensional plane (as a Schlegel diagram, for example [22]). This, in turn, means that the graph representation of every canonical capsid is expected to be planar, whether or not the three-dimensional canonical capsid is convex. From this it follows that the graphs describing connectivity for canonical capsids must abide by Euler's polyhedral formula.

Euler's polyhedral formula shows that various elements within convex and certain nonconvex polyhedra may be related to each other in a predictable manner. Specifically, the number of vertices (V), edges (E), and faces (F) of such a polyhedron may be related through the following equation:

$$V - E + F = 2. \tag{1}$$

All tilings describing the canonical capsid must satisfy this equation. This offers a direct relationship between the number of subunits (F) and the allowable number of interactions or edges per tile (σ). We will be able to relate the two terms (F and σ) by relating V, E, and F to the number of sides (σ), and the size of the capsid (T). From the general definitions of quasiequivalent spherical virus capsids (discussed above), we will establish the relationships between V, E, F, σ, and T to establish an expression for σ.

Since we know that the number of faces corresponds to the number of subunits in the canonical capsid, we have

$$F = 60T. \tag{2}$$

Also, as this is an edge-edge tiling, each edge is shared by exactly two faces. So the number of edges is

$$E = F\sigma/2 = 30T\sigma. \tag{3}$$

The Caspar and Klug rules of capsid assemblies shows that there are 12 five-valent vertices and $10(T-1)$ six-valent vertices [1] within the canonical capsid (we call these the centroid vertices, which are V_C in number). So the number of centroid vertices

$$V_C = 10(T-1) + 12 = 10T + 2. \tag{4}$$

We know that each prototile must have exactly one centroid vertex. Also, from our earlier definition of the prototile, the rest of the $(\sigma-1)$ vertices are trivalent in nature. Therefore, the number of trivalent vertices are equal to

$$V_R = 60T(\sigma-1)/3 = 20T(\sigma-1). \tag{5}$$

Finally, from Eqs. (4) and (5), we have the total number of vertices

$$V = V_R + V_C = 2 + 20T\sigma - 10T. \tag{6}$$

Substituting Eqs. (6), (3), and (2) in Eq. (1), we get

$$(2 + 20T\sigma - 10T) - (30T\sigma) + 60T = 2. \tag{7}$$

Further reducing this equation, the T's cancel out and we get

$$\sigma = 5. \tag{8}$$

Via simple topological relationships, we obtained an invariant geometric characteristic (σ) that is independent of the T number. Using the assumptions of a fixed number of faces, quasiequivalence, and trivalency for nonquasiequivalent vertices (i.e., modeling a 5/6,3-vertex polyhedron with $60T$ faces), topology dictates that only a five sided prototile will be able to assemble into any sized (T-numbered) canonical capsid (with the modification of subunit-subunit interaction angles). The number of sides (σ) is invariant of the triangulation (T) number, which shows that a subunit may be modified to assemble into capsids of non-native T numbers.

2. The shapes available to a five-sided canonical capsid subunit

As per the definition of the canonical capsid, any prototile that can tile a bound canonical capsid should, by the rules of quasiequivalence, also tile a two-dimensional hexavalent lattice (a hexagonal lattice containing hexavalent vertices) [1,21]. So our second query is simplified: How many five-sided prototiles may assemble into a hexavalent lattice?

The tiling of 2D surfaces by convex pentagons has been a subject studied at least since the early 20th century. In the early 1980s, Grünbaum and Shepherd enumerated a list of 13 known convex pentagonal edge-edge tilings, and, since then, one more has been added to that list [23,24]. We searched the

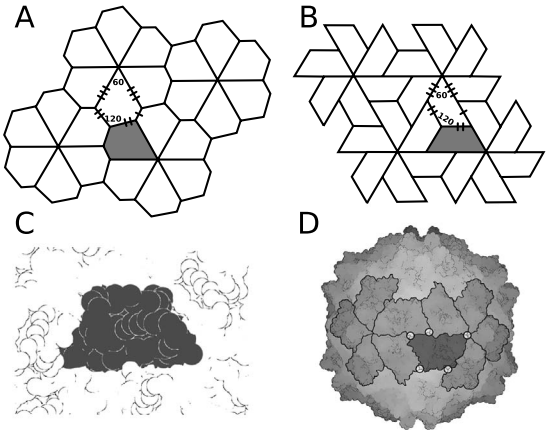

FIG. 5. Pentagonal type 5 tilings. Type 5 tiles [e.g., (a) and (b)] are the only known five-sided tiles that possess six-valent vertices. The interacting edges of the prototile in both (a) and (b) have been marked with one, two, and three dashes. Such edges interact with identically labeled edges of neighboring tiles to form interfaces within dimers, trimers, and hexamers, respectively. Interestingly, a version of this tile (b) resembles a very commonly seen subunit shape found in nature depicted as a space-filled subunit (c) and a subunit within its capsid environment (d) (structure used: 1vak of the sobemoviridae family).

resulting catalogs as listed by Sugimoto and Ogawa [23] for convex pentagonal prototiles that can assemble into a hexavalent lattice, and found only one tiling to fit that criterion [known in [23] as type 5; shown here in Fig. 5(a)]. Therefore, only one type of tiling among the known ones (the type 5 tiling) will be allowed to form a canonical capsid of any size.

D. Further implications

1. The ubiquitous trapezoid shape emerges

Interestingly, this tiling is equivalent (combinatorially) to the tiling seen in Fig. 5(b), which represents a 2D projection of the trapezoid structure seen in nature. This subunit shape is widely distributed in the spherical capsid world [25] and is found in virus families that differ greatly in sequence space (down to 0% sequence identity), host specificity (insects, birds, mammals), and capsid size (with recorded T numbers ranging from 1 through 13);[2] An example of one of these subunits is shown in Figs. 5(c) and 5(d). This suggests that topological constraints may play dominant roles in the natural design of the shape of virus capsid subunits. This, in turn, means that the chance that two proteins may have evolved from distinct lineages into the same structural shape cannot be dismissed.

2. Rationalization of the need for a five-sided subunit

One could look at the need for a five-sided subunit intuitively from a symmetry-oriented standpoint. There are two

[2]Note that, although the trapezoid *shape* is four sided, in the context of the capsid arrangement, the long edge interacts with two other subunits. Thus, graph theoretically, a vertex must bisect that long edge, giving us a bisected trapezoid.

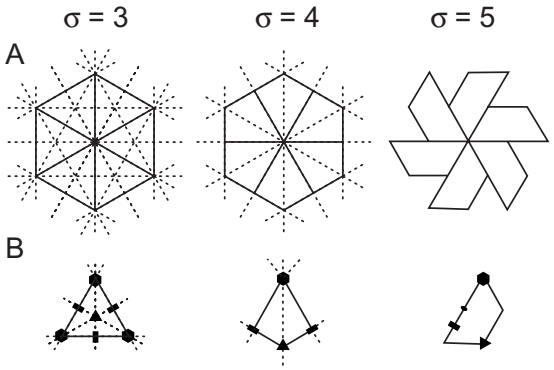

FIG. 6. Relationship between mirror symmetry elements and the number of edges (σ) per hexavalent tile. This image illustrates trigonal, tetragonal, and pentagonal tiles in (a) the hexavalent form and (b) as a single tile (along with the two-, three-, and sixfold symmetry relationships denoted). The dashed lines in the figure indicate possible mirror planes cutting the plane of the paper perpendicularly. It is immediately evident that the hexameric cluster for only $\sigma=5$ has no possible mirror symmetries, which is crucial in the chiral-centric biological world.

kinds of icosahedral point group symmetries—*full* (or achiral) icosahedral symmetry and *rotational* (or chiral) icosahedral symmetry. Although both point group symmetries have the same number of two-, three-, and fivefold axes of rotation, they differ in the number of mirror planes they possess; the full icosahedral symmetry group possesses 15 mirror planes while the rotational icosahedral symmetry group possesses none. Due to the chirality of biological macromolecules, it follows that only the chiral icosahedral symmetry may exist.

Figure 6 represents versions of the three-, four-, and five-sided prototile in its hexameric form [Fig. 6(a)], which is a precursor to the canonical capsid, and the prototile itself [Fig. 6(b)]. It is immediately evident that the number of possible mirror symmetries (shown as dashed lines) diminishes to zero at $\sigma=5$. As chirality (or the absence of mirror symmetries) is crucial within the biological world, $\sigma=5$ is the only viable option for virus capsid subunits; and from $\sigma=5$, the trapezoid of the tiling is automatically obtained.

3. Canonical capsids do represent group 1 capsids

All capsids present within group 1 families described in Table I—the capsids representable by monohedral tilings—also possess trapezoidal shapes. This means that the predictions made by analyzing canonical capsids, and the empirical evidence obtained from the shapes of group 1 capsids indicate a congruence between the mathematical models and natural capsids. Thus, the canonical capsid may, in fact, be used to represent and study capsids belonging to group 1 families of Table I. Furthermore, it is interesting to note that capsids belonging to families such as picornaviridae and comoviridae that possess chemically *distinct* subunits in the capsid asymmetric unit, but that appear to have no gross structural overlaps and holes, also possess the familiar trapezoidal subunit motif [26].

Capsids that possess subunit-subunit holes, overlaps, or structural variability (i.e., group 2 capsids) were found to

possess, largely, nontrapezoidal-shaped subunits. Most notable among these are the polyomaviridae family viruses [27,28] that not only break the rules of monohedral tilability, but also break the otherwise longstanding rules of quasiequivalence. We propose that the breaking of Caspar and Klug's rules of quasiequivalence [1] is possible only with the use of (1) chemically distinct subunits or (2) subunits that grossly break the rules of monohedral tilability.

4. Divergence or convergence?

The trapezoidal subunit shape is seen within capsids sharing little to no similarity in amino acid sequence, host specificity, genome type (RNA, DNA), and size (*T* number). Our findings raise interesting questions about the evolutionary reason for this structure's prevalence among such diverse virus families. It is widely believed that all these capsids arose from a single protoviral strain [25]. However, we show that the bisected trapezoid is the only subunit shape (among other type 5 tilings) that is allowable within canonical capsids, i.e., it appears as though the trapezoid is the only shape available to a capsid that attempts to maximize, via edge-edge tilings, the amount of interactions, while minimizing the extent of holes, and simultaneously reducing design complexity by being relatively structurally invariant or monohedral (all of these features are acceptable from a free energetic point of view). Consequently, it is not improbable that distinct families evolved in parallel and encountered, independently, the trapezoidal-shaped jelly-roll architecture, i.e., convergent evolution is a distinct possibility, especially in systems with general but stringent constraints (such as topological constraints).

5. Future applications

Virus tiling theory, as pioneered by Twarock, has already been used to explain interesting assembly-related properties of capsids that break the rules of quasiequivalence by displaying only pentamers and no hexamers [12,29–31]. Specifically, they showed that capsids belonging to the polyomaviridae family must be represented as bound tilings formed from *two* distinct subunit shapes or tiles [12,29], i.e., capsids from this family cannot be represented by monohedral tilings, which corroborates the classification of these capsids as group 2 in Table I. Importantly, Twarock and colleagues showed that one may characterize the "assembly pathways" of these tilings using mathematical methods [30,31]. Those studies were primarily applied to capsids belonging to the polyomaviridae family.

Our findings show that canonical capsids—monohedral bound trapezoidal tilings that follow the rules of quasiequivalence—may be used to represent a large number of capsids, allowing for a physical understanding of those capsids in a manner that builds upon the techniques introduced by Twarock and colleagues. Mathematical and physical investigations of these canonical capsids are currently being pursued.

IV. CONCLUSION

We have shown that a majority of capsids (comprising eight out of twelve structurally characterized families) may be represented as monohedral tilings (canonical capsids), thereby allowing the use of many fields of mathematics (such as topology, tiling theory, graph theory, etc.) to explain capsid structure and dynamics. We demonstrated that only a specific five-sided prototile is compatible with bound monohedral tilability and the definitions of quasiequivalence of virus capsids. This we did using simple topology, graph theory, and symmetry concepts. Of the possible tilings, one—of trapezoidal shape—was found to be displayed by capsid subunits within every one of the eight tilable virus capsid families. This trapezoidal subunit shape is found in capsids sharing little to no sequence similarity, host specificity, genome type (RNA, DNA), and size (*T* number). These results raise interesting questions about the nature of the noncanonical capsids, and the divergent and convergent nature of the group 1 capsids with respect to evolution and design. More importantly, our result that the expected shape of the canonical capsid accurately reflects the shape of the group 1 capsid subunits found in nature lends credence to usefulness of the bound monohedral tiling model of capsids (the "canonical capsid").

ACKNOWLEDGMENTS

We thank Professor Jack Johnson and Professor Vijay Reddy for their comments and discussion. This work was supported by the National Institutes of Health (Grant No. RR012255) and the National Science Foundation (Grant No. PHY0216576).

[1] D. L. D. Caspar and A. Klug, Cold Spring Harbor Symp. Quant. Biol. **27**, 1 (1962).
[2] K. Gopinath, S. Sundareshan, M. Bhuvaneswari, A. Karande, M. R. Murthy, M. V. Nayudu, and H. S. Savithri, Indian J. Biochem. Biophys. **31**, 322 (1994).
[3] F. Coulibaly, C. Chevalier, I. Gutsche, J. Pous, J. Navaza, S. Bressanelli, B. Delmas, and F. A. Rey, Cell **120**, 761 (2005).
[4] P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten, Structure **14**, 437 (2006).
[5] A. Arkhipov, P. Freddolino, and K. Schulten, Structure **14**, 1767 (2006).
[6] R. Zandi, D. Reguera, R. F. Bruinsma, W. M. Gelbart, and J. Rudnick, Proc. Natl. Acad. Sci. U.S.A. **101**, 15556 (2004).
[7] R. Zandi and D. Reguera, Phys. Rev. E **72**, 021917 (2005).
[8] T. Chen, Z. Zhang, and S. C. Glotzer, Proc. Natl. Acad. Sci. U.S.A. **104**, 717 (2007).
[9] K. Van Workum and J. F. Douglas, Phys. Rev. E **73**, 031502 (2006).
[10] D. C. Rapaport, Phys. Rev. E **70**, 051905 (2004).
[11] H. D. Nguyen, V. S. Reddy, and C. L. Brooks III, Nano Lett.

7, 338 (2007).
[12] R. Twarock, Philos. Trans. R. Soc. London, Ser. A **364**, 3357 (2006).
[13] R. S. Schwartz, R. L. Garcea, and B. Berger, Virology **268**, 461 (2000).
[14] M. F. Hagan and D. Chandler, Biophys. J. **91**, 42 (2006).
[15] D. Endres, M. Miyahara, P. Moisant, and A. Zlotnick, Protein Sci. **14**, 1518 (2005).
[16] N. Destainville, M. Widom, R. Mosseri, and F. Bailly, J. Stat. Phys. **120**, 799 (2005).
[17] C. M. Shepherd, I. A. Borelli, G. Lander, P. Natarajan, V. Siddavanahalli, C. Bajaj, J. E. Johnson, C. L. Brooks III, and V. S. Reddy, Nucleic Acids Res. **34**, D386 (2006).
[18] B. Grünbaum and G. C. Shephard, *Tilings and Patterns*, 2nd ed. (W. H. Freeman, New York, 1986).
[19] B. Grünbaum and G. C. Shephard, Bull. Am. Math. Soc. **3**, 951 (1980).
[20] J. E. Johnson and J. A. Speir, J. Mol. Biol. **269**, 665 (1997).
[21] J. F. Sadoc and R. Mosseri, *Geometrical Frustration* (Cambridge University Press, Cambridge, U.K., 1999).
[22] P. W. Fowler and K. M. Rogers, J. Chem. Inf. Comput. Sci. **41**, 108 (2001).
[23] T. Sugimoto and T. Ogawa, Forma **15**, 75 (2000).
[24] T. Sugimoto and T. Ogawa, Forma **20**, 1 (2005).
[25] M. G. Rossmann and J. E. Johnson, Annu. Rev. Biochem. **58**, 533 (1989).
[26] T. Lin and J. E. Johnson, Adv. Virus Res. **62**, 167 (2003).
[27] T. Stehle, S. J. Gamblin, Y. Yan, and S. C. Harrison, Structure **4**, 165 (1996).
[28] R. C. Liddington, Y. Yan, J. Moulai, R. Sahli, T. L. Benjamin, and S. C. Harrison, Nature (London) **354**, 278 (1991).
[29] R. Twarock, J. Theor. Biol. **226**, 477 (2004).
[30] T. Keef, A. Taormina, and R. Twarock, Phys. Biol. **2**, 175 (2005).
[31] T. Keef, C. Micheletti, and R. Twarock, J. Theor. Biol. **242**, 713 (2006).