

The influence of symmetry on the probability of assembly pathways for icosahedral viral shells

Miklós Bóna^{a1} and Meera Sitharam^{b*2}

^aDepartment of Mathematics, University of Florida, Gainesville, FL, USA; ^bCISE Department, University of Florida, Gainesville, FL, USA

(Received 19 March 2008; final version received 7 May 2008)

This paper motivates and sets up the mathematical framework for a new program of investigation: to isolate and clarify the precise influence of symmetry on the probability space of assembly pathways that successfully lead to icosahedral viral shells. Several tractable open questions are posed. Besides its virology motivation, the topic is of independent mathematical interest for studying constructions of symmetric polyhedra. Preliminary results are presented: a natural, structural classification of subsets of facets of $T = 1$ polyhedra, based on their stabilizing subgroups of the icosahedral group; and a theorem that uses symmetry to formalize why increasing depth increases the numeracy (and hence probability) of an assembly pathway type (or symmetry class) for a $T = 1$ viral shell.

Keywords: icosahedral group; viral shell; polyhedral graph; probability; assembly pathways

1. Introduction

The viral assembly process – just like many other spontaneous macromolecular assembly processes – is not well-understood. This lack of understanding prevails even for basic assembly of $T = 1$ viral shells, including those that assemble with equal efficacy into empty shells without enclosing the internal genomic material, and without the use of chaperone or scaffolding proteins. This is the type of assembly that we consider here. Many mathematical models of viral shell assembly have been proposed and studied including [2,3,9,11–13,18–20]. Here, we use the viral assembly pathway *GT* model of Ref. [16] that was developed to answer focused questions that *concern only* the influence of symmetry and geometric constraints (present in the complete viral shell) on the relative probabilities of different types of abstract assembly pathways. The geometric constraints (within monomers or between monomers) are extracted from the X-ray or cryo-EM structure of the viral shell. More specifically, the final viral structure can be viewed formally as the solution to a symmetric system of geometric constraints that can be expressed as algebraic equations and inequalities.

The basic *GT* model does not directly address factors resulting from dynamics or kinetics or solvent interaction, although it permits a natural extension to incorporate such factors.

As a result of this simplicity, the basic *GT* model is easily tunable and provides an intuitive theory of assembly that is restricted to (and based purely on) geometry and symmetry constraints present in the complete viral shell. Moreover, the *GT* model is computationally tractable, *i.e.* there is an efficient randomized algorithm for computing (a provably good approximation of) probability distribution on pathways. Since the required algorithms are modifications of state-of-the-art 3D geometric constraint decomposition algorithms [15], simulation software for the

*Corresponding author. Email: sitharam@cise.ufl.edu

model is built directly upon existing opensource software for 3D geometric constraint solving [14]. For biochemical validation, a precise GT model for minute virus of mice (MVM) assembly pathways has been developed using MVM X-ray structure's geometric constraints, and rough models have been developed for maize streak geminivirus (MSV), and human adeno-associated virus (AAV4). Several of the GT model's predictions consistently explain existing experimental observations about $T = 1$ viral shell assembly pathways for MVM, MSV and AAV4. Furthermore, the GT model's biochemical and mathematical assumptions have been clarified and justified.

2. Motivation and contribution

In Ref. [17], the authors observed the following. The study of the probability of an abstract pathway is reduced to two separate factors – geometry and symmetry – that could largely be treated independently. We explain briefly below. The GT model's abstract *pathways* represent static information in the construction (or decomposition) of the complete viral shell (Figure 1). The nodes of these pathway trees represent subassemblies that do not disintegrate during the course of the assembly process.

See Figures 2 and 3 for examples of pathways. These subassemblies are subsystems of the complete viral geometric constraint system. In a pathway tree, these subassemblies or subsystems are partially ordered by containment, with the root representing the complete assembled structure, *i.e.* we only consider trees that represent *successful* pathways. Additionally, in a simplified pathway tree, the leaves represent the individual viral shell monomers. Since the $T = m$ viral shell can be overlaid on a polyhedron with the monomers represented as facets, a simplified pathway tree for an icosahedral $T = m$ viral shell describes the assembly of a corresponding polyhedron from its individual facets. This manuscript will only refer to simplified, successful pathways. See Figure 1 for nomenclature of a $T = 1$ polyhedron that we use in this paper.

In the GT model, the first factor influencing the probability of a given pathway tree is the *geometric stability factor*: this is determined by quantifiable properties – such as extent of rigidity, algebraic complexity and size of configuration space – for each of the subassemblies (internal nodes) that appear in the pathway tree. This factor is estimated by analysing the corresponding subsystems of the given viral geometric constraint system. It was shown

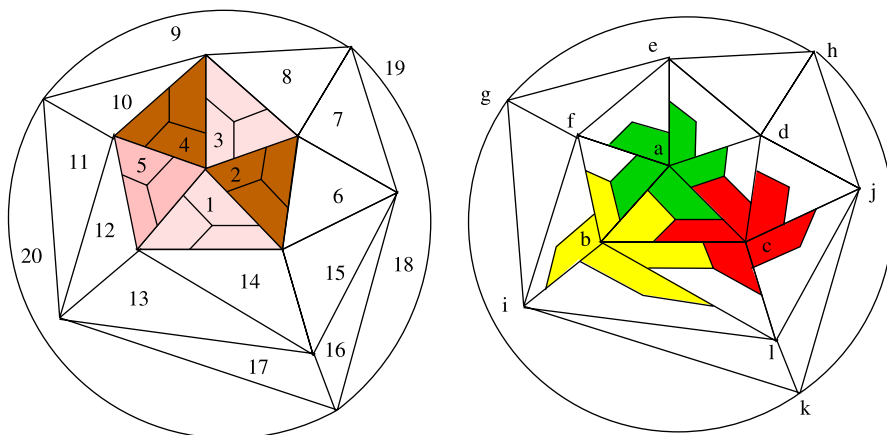


Figure 1. (Left) Facenumbers: pentamer of trimers in a $T = 1$ polyhedron. (Right) Vertex numbers: trimers of pentamers in a $T = 1$ polyhedron. Available in colour online.

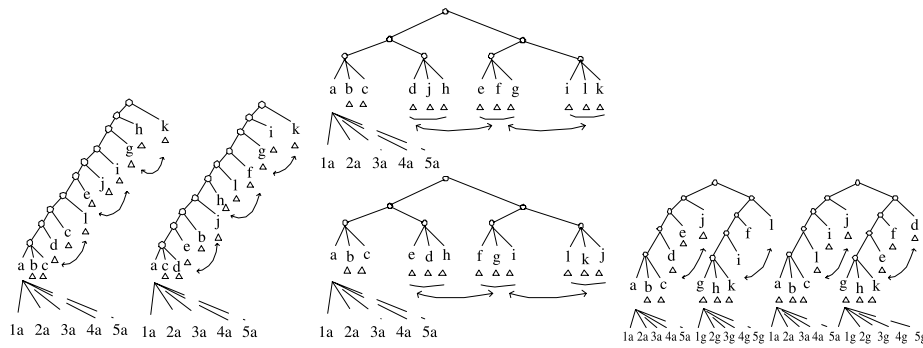


Figure 2. $T = 1$ assembly pathways based on pentameric subassemblies and nomenclature of Figure 1, triangles at bottom represent pentamers; arrows represent the action of the icosahedral group on pathways; only the two arrows to the right of the two middle figures fix the corresponding pathways.

in Ref. [16] that the geometric stability factor can be expressed purely graph theoretically, using the geometric constraint graph underlying the viral constraint system. This factor is correlated with biochemical stability influenced by assembly and dis-assembly energy thresholds. Some pathways can never occur (have probability zero) since the subassemblies occurring in them are unstable (their geometric stability factor is zero). Such pathways are *geometrically invalid*. It was further shown in Ref. [17] that validity of pathways can be expressed using even simpler polyhedral graphs extracted from the underlying geometric constraint graphs, often directly using the $T = m$ polyhedral graphs.

The second factor that clearly influences the probability of a given symmetry type or class of pathway τ is the size of the orbit of τ under the action of the icosahedral group (Figure 2). By a classic result [1,10], the sense-preserving automorphism group of a polyhedral graph can be realized as the group of rotations of a convex polyhedron whose skeleton is the given polyhedral graph. Hence, we can continue to use a purely graph-theoretic treatment of subassemblies and valid pathways when analysing this factor as well. In this paper, we study this *symmetric numeracy factor* exclusively.

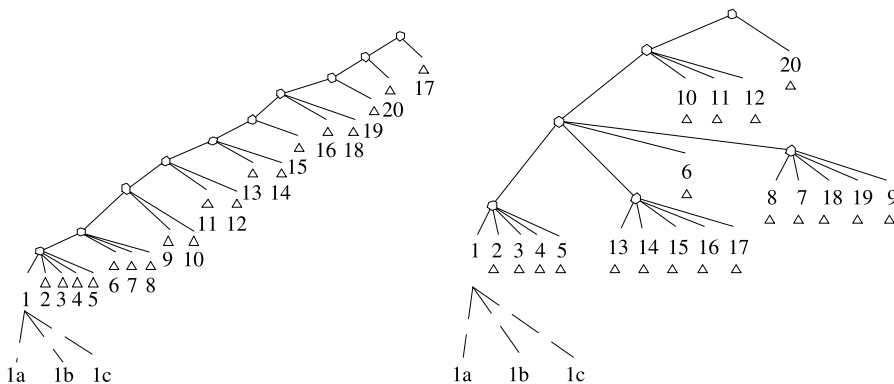


Figure 3. $T = 1$ assembly pathways based on trimeric subassemblies, triangles at bottom represent trimers.

Remark. Note that the geometric stability factor and validity remain invariant for any pathway tree of the same *symmetry type*, or in the same *symmetry class* (using the symmetries of the underlying polyhedron or the automorphisms of the corresponding polyhedral graph).

2.1 Contribution and organisation

With the above virology motivation, the overall contribution and organization of this paper is as follows.

(1) In Section 3, we set up the required definitions with examples and develop a program of questions concerning assembly pathway trees τ of polyhedra or polyhedral graphs P with a symmetry/automorphism group G .

This is a novel approach to studying symmetric macromolecular assembly, isolating and clarifying the precise influence of the symmetric numeracy factor on the probability of assembly pathway types. While the questions posed here stem from the above virology motivation, to the best of our knowledge, they begin a new direction of mathematical investigation.

Related work. There is a rich literature on the enumeration of construction sequences of symmetric polyhedra and their underlying polyhedral graphs, including fullerenes and fulleroids [4,6,7]. These have focused on enumerating (constructions of) *different* polyhedra, using structural features such as the so-called *zigzags*. On the one hand, our goal – of classifying pathways based on their stabilizers – is geared towards enumerating (symmetry classes) of constructions of the *same* polyhedron representing the viral shell. Hence, while we expect to draw on the existing work on zigzags, we believe that the questions and theorems discussed here require new ingredients.

(2) In Section 4, the tractability of these questions is illustrated by new results about assembly pathway trees of icosahedral $T = 1$ polyhedra P . In Theorem 1, we observe that the probability that P assembles using a pathway tree τ generally increases with the depth of τ : more precisely, the symmetric numeracy factor, *i.e.* the size of the orbit (or symmetry class) of τ , is bounded below by the depth of τ . This is a useful theorem that directly indicates that the symmetric numeracy factor generally increases with the depth. Moreover, increasing depth also means that each intermediate subassembly or node in the pathway tree is assembled from fewer constituent subassemblies (its children); and it is intuitively clear that this would in turn effect an increase in the geometric stability factor of the pathway (this is formalized in Ref. [16]). Hence, Theorem 1 clarifies how increasing depth increases the probability of the pathway symmetry type in two independent ways. The proof of Theorem 1 uses a structural classification: for each subgroup S of the icosahedral group, we describe the structure of a subset of P 's facets that is fixed by S .

3. Framework and questions

A $T = 1$ polyhedral graph is the dual or facet-adjacency graph of Figure 1 (where each of the 20 faces of the icosahedron is subdivided into three 5-sided facets as shown). That is, this graph has one vertex for each facet shown in the figure, and edges connect vertices corresponding to pairs of adjacent facets. As mentioned earlier, we will identify the $T = 1$ polyhedral graph P with any polyhedron whose facet adjacencies are specified by P and whose symmetry group is the sense-preserving automorphism group of P . We will use these objects interchangeably. We call such a polyhedron (with 60 facets) a $T = 1$ polyhedron, *i.e.* it is based on an icosahedral construction with T number 1 (see [5] for Caspar–Klug T number or alternatively [8] for the equivalent Goldberg–Coxeter (h, k) -numbers). A $T = 1$ assembly pathway tree is a rooted tree whose leaves are labelled by the 60 facets of a $T = 1$ polyhedron (or the 60 vertices of a $T = 1$ polyhedral graph). Each internal node is labelled by the subset of facets corresponding to the

descendant leaves of that node (and the corresponding subgraph of the $T = 1$ polyhedral graph), *i.e.* it represents the subassembly induced by them. The root represents the complete successful assembly of a $T = 1$ polyhedron or the entire $T = 1$ polyhedral graph. From now on, a *pathway tree* refers to a $T = 1$ pathway tree. The *depth* of a pathway tree is the number of edges on the longest path from root to any leaf.

It is well-known that if G is a finite permutation group acting on a set S and $s \in S$, then $|s^G| = |G|/|G_s|$, where s^G is the *orbit* or *symmetry class* of s under the action of G , and G_s is the *stabilizer subgroup* of G that fixes s . Here, $|\cdot|$ denotes size or cardinality of a set. If $U \subseteq S$, then the stabilizer G_U of U is also a subgroup of G .

We will study the natural action of the icosahedral group on the set of facets of a $T = 1$ polyhedron, set of subsets of facets and the set of pathway trees (induced by the action on the facets labelling the leaves). A property \mathcal{Q} of pathway trees τ is *symmetry-invariant* provided $\mathcal{Q}(\tau')$ either holds for all τ' in τ 's symmetry class, or holds for none of them. Note that a conjunction of symmetry-invariant properties is also symmetry invariant. The following is an example of a symmetry-invariant property whose probability of occurrence would be valuable to estimate. Pathway trees with this property can be biochemically interpreted as assembly pathways that are based on a specific type of *nucleation*.

Example [nucleation]. A *trimer* is a triangle subgraph (3-cycle) of a $T = 1$ polyhedral graph induced by any three facets of a $T = 1$ polyhedron that cover one face of the underlying icosahedron. Similarly, a *pentamer* is a pentagon subgraph (5-cycle) induced by any five facets of a $T = 1$ polyhedron that are incident at any of the vertices of the underlying icosahedron. A *dimer* is an edge between any two adjacent facets belonging, respectively, to two adjacent faces of the underlying icosahedron. A symmetry-invariant property of pathway trees is whether it is a pentamer nucleation tree (respectively dimer or trimer nucleation tree), *i.e.* all nodes whose children are all leaves, correspond to pentamers (respectively dimers, trimers).

As remarked earlier, any natural notion of *validity* of pathways is a symmetry-invariant property and Ref. [17] shows how some natural notions of validity can be expressed directly using the $T = 1$ polyhedral graphs. Below are two such examples of *validity*.

Example [validity]. A node in a pathway tree is valid if the subgraph of the $T = 1$ polyhedral graph corresponding to it is connected. Stronger notions of validity are defined by specifying a set of (symmetry classes of) *base stable* subgraphs of the $T = 1$ polyhedral graph [17]. In this case, a node ν in a pathway tree is valid if one of the base stable subgraphs has a non-trivial intersection with the subgraphs associated with each of ν 's children, where 'non-trivial' is defined appropriately. In all cases, a pathway tree is valid if all its nodes are valid.

With these definitions and examples, we can pose the following natural questions (the optional and variable contents in the parentheses give many versions of each question).

Question 1: Given a subgroup G of the icosahedral group, characterize the set of pathway trees (respectively with a given symmetry-invariant property, say validity or nucleation or both) stabilized by that subgroup. Alternatively, given a pathway tree (respectively with a given symmetry-invariant property) specify the information necessary and sufficient to determine its stabilizer.

Question 2: Count or enumerate (respectively symmetry classes of) pathway trees (respectively with a given symmetry-invariant property).

Question 3: Additionally, it is interesting to compute the ratio of the number of pathway trees (respectively number of symmetry classes) that satisfy two symmetry-invariant properties \mathcal{Q}_1

(say nucleation) and \mathcal{Q}_2 (say validity), to the number of pathway trees (respectively number of symmetry classes) that satisfy only \mathcal{Q}_2 (validity).

Question 4: The above questions directly extend to larger icosahedrally symmetric polyhedra (larger T numbers), to fullerenes and fulleroids and polyhedra with different symmetry groups. In such cases, the questions can also be phrased as algorithmic questions, where asymptotic complexity of the algorithm is expressed in terms of the number of facets of the polyhedron (or the T number).

4. Results

It is well-known that the icosahedral group A_5 has the following subgroups and subgroup sizes: A_5 , of size 60; A_4 , of size 12; D_5 , of size 10; D_3 , of size 6; Z_5 , of size 5; $Z_2 \times Z_2$, of size 4; Z_3 , of size 3; Z_2 , of size 2; and id , of size 1.

The group A_5 acts on the set of 60 facets of a $T = 1$ polyhedron in a natural way. So the possible symmetry class sizes are 60, 30, 20, 15, 12, 10, 6, 5 and 1. It is not difficult to verify that all these class sizes actually occur.

Let S be a subset of the set of facets of a $T = 1$ polyhedron, and assume that S is fixed by a subgroup H of A_5 . Then we have the following possibilities for S . (These can be verified by considering how each subgroup of A_5 acts on the facets of the $T = 1$ polyhedron.)

- (1) If $H = A_5$, then S is the set of all 60 facets.
- (2) If $H = A_4$, then S consists of all the facets situated on four full icosahedral faces whose centres form a regular tetrahedron.
- (3) If $H = D_5$, then S is somewhat more complicated to describe. In this case, H fixes a set of two pentagons in parallel planes whose edges are edges of the icosahedron. Then S can be any set of facets arranged symmetrically around these ten edges. (So for instance, the number of facets in S is always divisible by 10 in this case).
- (4) If $H = D_3$, then S consists of two full opposite faces of the icosahedron, and possibly, additional facets added on each of the six sides of these faces in a symmetrical fashion.
- (5) If $H = Z_5$, then S consists of 5-tuples of facets arranged symmetrically around the same axis of rotation.
- (6) If $H = Z_2 \times Z_2$, then S consists of facets adjacent to two opposite edges of the icosahedron.
- (7) If $H = Z_3$, then S is a union of triples of facets, the triples consisting of facets symmetrically arranged around the same axis of rotation.
- (8) If $H = Z_2$, then S consists of pairs of opposite facets (on opposite faces of the icosahedron).
- (9) If $H = 1$, then S can be any subset of facets not belonging to any of the previous eight cases.

We are now ready to state and prove a theorem on the sizes of icosahedral symmetry classes of $T = 1$ pathway trees.

THEOREM 1. Let $d = 3$. Then the size of the symmetry class of a $T = 1$ pathway tree τ is at least $d - 1$, where d is the depth of τ .

Proof. The proof is by verifying all cases depending on the values of d . In each case, we will use the description of S given in the list above. Let G be the icosahedral group. We consider the longest path in τ (of length at least d). At each level, consider the set of leaves in τ of that depth. Let S denote the smallest such set; and let $Stab_S$ denote its stabilizer.

- (1) If $d > 30$, then by the Pigeon-hole Principle, there is a level in τ where S consists of only one leaf. This can be any of the 60 facets of the $T = 1$ polyhedron, so τ has a symmetry class size 60.
- (2) If $20 < d \leq 30$, then again by the Pigeon-hole Principle, there is a level in τ where S consists of at most two leaves. If they correspond to opposite facets, then there are 30 possibilities for their pairing, otherwise there are more. So τ has a symmetry class size of at least 30.
- (3) If $15 < d \leq 20$, then there is a level in τ for which S consists of at most three leaves. (If there is a level with two leaves, we are in the previous case. For this reason, in all subsequent cases, we will assume that the smallest level is *not* smaller than it has to be.) Then $Stab_S$ can consist of at most three elements – that happens when $Stab_S = Z_3$), so the orbit S^G of S is of size at least 20.
- (4) If $12 < d \leq 15$, then there is a level in τ that consists of at most four leaves. The above description shows that if S has four elements, then $|Stab_S| \leq 4$, with equality holding when $Stab_S = Z_2 \times Z_2$. Therefore, $|S^G| \geq 15$.
- (5) If $10 < d \leq 12$, then there is a level in τ that consists of at most five leaves. When this is exactly five, $|Stab_S| \leq 5$ (equality occurs when $Stab_S = Z_5$). Therefore, $|S^G| \geq 12$.
- (6) If $6 < d \leq 10$, then we take S to be the minimum number of *nodes* at any given level and show the stronger statement that $|S^G| \geq 10$. Indeed, the only way for this claim to be false would be if S were stabilized by a subgroup of G that has more than six elements, that is, by A_5 , A_4 or D_5 . The list above shows that that is impossible since $|S|$ has less than 10 elements.
- (7) If $d = 6$, then there is a level in τ that consists of at most 10 nodes. When $|S| = 10$, then it is possible that $|Stab_S| = 10$, provided that $Stab_S = D_5$, but the larger subgroups of A_5 , that is, A_5 and A_4 , cannot be stabilizers of S . Therefore, $|S^G| \geq 6$.
- (8) If $4 < d \leq 5$, the stabilizer of S is a *proper* subgroup of A_5 , and as such has at most 12 elements. Therefore, the orbit of S is of size at least 5.

Notes

1. Supported in part by NSF grant DMS0714912. Email: bona@math.ufl.edu
2. Supported in part by NSF grants NER0404116, DMS0714912.

References

- [1] L. Babai and W. Imrich, *Sense preserving groups of polyhedral graphs*, Monatsh. Math. 79 (1975), pp. 1–2.
- [2] B. Berger, P. Shor, J. King, D. Muir, R. Schwartz, and L. Tucker-Kellogg, *Local rule-based theory of virus shell assembly*, Proc. Natl Acad. Sci. USA 91 (1994), pp. 7732–7736.
- [3] B. Berger and P.W. Shor, *Local rules switching mechanism for viral shell geometry*, Tech. Rep. MIT-LCS-TM-527 (1995).
- [4] G. Brinkmann and A. Dress, *A constructive enumeration of fullerenes*, J. Algorithms 23 (1997), pp. 345–358.
- [5] D. Caspar and A. Klug, *Physical principles in the construction of regular viruses*, Cold Spring Harb. Symp. Quant. Biol. 27 (1962), pp. 1–24.
- [6] M. Deza and M. Dutour, *Zigzag structures of simple two-faced polyhedra*, Combin. Probab. Comput. 14(1–2) (2005), pp. 31–57.
- [7] M. Deza, M. Dutour, and P.W. Fowler, *Zigzags, railroads, and knots in fullerenes*, Chem. Inf. Comp. Sci. 44 (2004), pp. 1282–1293.
- [8] M. Goldberg, *A class of multisymmetric polyhedra*, Tohoku Math. J. 43 (1937), pp. 104–108.
- [9] J.E. Johnson and J.A. Speir, *Quasi-equivalent viruses: A paradigm for protein assemblies*, J. Mol. Biol. 269 (1997), pp. 665–675.
- [10] P. Mani, *Automorphismen von polyhedrische graphen*, Math. Ann. 192 (1971), pp. 279–303.

- [11] C.J. Marzec and L.A. Day, *Pattern formation in icosahedral virus capsids: The papova viruses and nudaurelia capensis β virus*, Biophysics 65 (1993), pp. 2559–2577.
- [12] D. Rapaport, J. Johnson, and J. Skolnick, *Supramolecular self-assembly: Molecular dynamics modeling of polyhedral shell formation*, Comp. Phys. Comm. 121–122 (1999), pp. 231–235.
- [13] V.S. Reddy, H.A. Giesing, R.T. Morton, A. Kumar, C.B. Post, C.L. Brooks, and J.E. Johnson, *Energetics of quasiequivalence: Computational analysis of protein–protein interactions in icosahedral viruses*, Biophysics 74 (1998), pp. 546–558.
- [14] M. Sitharam, *Frontier, opensource gnu geometric constraint solver: Version 1 (2001) for general 2d systems; version 2 (2002) for 2d and some 3d systems; version 3 (2003) for general 2d and 3d systems*, Software available at <http://www.cise.ufl.edu/~sitharam>, <http://www.gnu.org> (2004).
- [15] ———, Combinatorial approaches to geometric constraint solving: problems, progress and directions. Geometric and algorithmic aspects of computer-aided design and manufacturing, *DIMACS Ser. Discrete. Math. Theoret. Comput. Sci.*, 67 (2005), pp. 117–163, Providence, RI: American Mathematical Society.
- [16] M. Sitharam and M. Agbandje-McKenna, *Modeling virus assembly using geometric constraints and tensegrity: Avoiding dynamics*, J. Comput. Biol. 13(6) (2006), pp. 1232–1265.
- [17] M. Sitharam and M. Bóna, *Combinatorial enumeration of macromolecular assembly pathways*, Proceedings of the International Conference on Bioinformatics and Applications World Scientific (2004).
- [18] A. Zlotnick, *To build a virus capsid: An equilibrium model of the self assembly of polyhedral protein complexes*, J. Mol. Biol. 241 (1994), pp. 59–67.
- [19] A. Zlotnick, R. Aldrich, J.M. Johnson, P. Ceres, and M.J. Young, *Mechanisms of capsid assembly for an icosahedral plant virus*, Virology 277 (2000), pp. 450–456.
- [20] A. Zlotnick, J.M. Johnson, P.W. Wingfield, S.J. Stahl, and D. Endres, *A theoretical model successfully identifies features of hepatitis b virus capsid assembly*, Biochemistry 38 (1999), pp. 14644–14652.