

# **Principles of Virus Capsid Design**

A thesis presented by

Ranjan V. Mannige

to

The Scripps Research Institute Graduate Program  
in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy  
in the subject of Biophysics

for

The Scripps Research Institute  
La Jolla, California

February 2010

© 2010 by Ranjan V. Mannige  
All Rights Reserved.

# THESIS ACCEPTANCE FORM

## The Scripps Research Institute

This thesis titled *Principles of Virus Capsid Design* by Ranjan V. Mannige has been read by each member of the following thesis committee and has been found to be satisfactory.

---

Date

---

Charles L. Brooks III, Thesis Advisor

---

Date

---

Elizabeth D. Getzoff, Committee Chair

---

Date

---

William M. Gelbart, External Member

---

Date

---

John S. Johnson

---

Date

---

Arthur J. Olson

---

Date

---

Vijay S. Reddy

## Acknowledgments

At the risk of sounding hackneyed, I want to first thank my parents, Tai, Ajja and Nagma for their support, and for their sometimes infuriating but often comforting “we believe in you”s. Then comes my mentor—Charles Brooks III. After staring at this screen for longer than is advised, I have concluded that there is no way of truly showing him my gratitude. I am without words. Thank you for everything. I also emphatically thank Elizabeth (Libby) Getzoff for being the most caring and coaching, motherly even, thesis committee chair I could ever hope for; discussions with Libby lifted onerous clouds of despair that one is often plagued with at dead ends and walls. I thank the rest of my committee—William Gelbart, Art Olson, Jack Johnson, and Vijay Reddy—for their always useful and multifaceted insights and seemingly inexhaustible patience. I thank Montgomery Pettitt for initiating my research trajectory by mentoring me in Houston. My aunt and uncle—Susan Paachi and Vikram Bappa—are thanked beyond any available gesture, since I would be nowhere without their guidance. They are my second parents.

I would like to thank Marylyn Rinaldi and Diane Kreger—the bastions of the Scripps Graduate School—for making me feel supported, even after my lab moved to Michigan. In Michigan itself, I want to thank my new partner-in-crime, Alana Canfield, for support and company, and for the caffeinated drinks. It’s hard to imagine life without my labmates past and present. I will miss them, especially my fellow East Wingers.

Captain Piyush Jadhav, thanks for checking up on me so often, and forcing me out of my research burrow with your calls. Shishir Nikam, thanks for giving me company (online) during my late nights. Shishir once claimed that my graduation will be possible only because of the Black Magic rituals that he performed for me every day. Don’t stop, Shishir. Don’t stop.

# Contents

<b>Title Page</b>	<b>i</b>
<b>Copyright notice</b>	<b>ii</b>
<b>Thesis Acceptance Form</b>	<b>iii</b>
<b>Acknowledgments</b>	<b>iv</b>
<b>List of Tables</b>	<b>ix</b>
<b>List of Figures</b>	<b>xi</b>
<b>Abstract</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Viruses: a general introduction . . . . .	1
1.2 Specific reasons to study the capsid . . . . .	3
1.3 Math and biology can be friends! . . . . .	5
1.3.1 Functional energy landscape . . . . .	5
1.3.2 Making the mathematical connection . . . . .	6
1.4 Limitations of math . . . . .	8
<b>2 Introduction II: Selected Topics</b>	<b>9</b>

2.1	Structural virology: then and now . . . . .	9
2.1.1	Milestone I: Equivalence and symmetry . . . . .	10
2.1.2	Milestone II: Scalable capsids (a confluence of disciplines) . . . . .	11
2.1.3	Making “scalable” capsids . . . . .	12
2.1.4	Quasi-equivalence in virus capsids . . . . .	14
2.1.5	Triangulation ( $T$ ) number . . . . .	14
2.1.6	Mini-summary . . . . .	15
2.2	Symmetry . . . . .	15
2.2.1	Rotational symmetry . . . . .	15
2.2.2	Platonic/High order structures . . . . .	16
2.2.3	Quasi-rotational symmetry . . . . .	18
2.2.4	Chiral subunit clusters can not possess mirror symmetries . . . . .	19
2.3	Introduction to nanotechnology . . . . .	20
2.4	Nanotechnology, history and viruses . . . . .	21
<b>3</b>	<b>Simplified Capsid Models Dictate Capsid Subunit Properties</b>	<b>24</b>
3.1	Part I. Modeling the spherical virus capsid . . . . .	25
3.1.1	The model . . . . .	26
3.1.1.1	Approximations are inevitable . . . . .	26
3.1.1.2	A plethora of models . . . . .	27
3.1.1.3	Introducing the canonical capsid . . . . .	28
3.1.2	Methods - I (Testing the existence of the canonical capsid) . . . . .	30
3.1.2.1	Virus capsids analyzed . . . . .	31
3.1.2.2	Test for capsid “tilability” . . . . .	31
3.1.2.3	Test for capsid “monohedrality” . . . . .	33
3.1.3	Tilability of natural spherical capsids . . . . .	34
3.2	Characterizing the subunit shape . . . . .	37

3.2.1	I. The number of sides ( $\sigma$ ) . . . . .	37
3.2.1.1	Representing the subunit . . . . .	38
3.2.1.2	Bringing out the math . . . . .	39
3.2.2	A capsid invariant . . . . .	41
3.2.3	II. The shapes available to a five-sided subunit . . . . .	41
3.2.4	The ubiquitous trapezoid shape emerges . . . . .	43
3.3	Divergence or convergence? . . . . .	44
3.4	Canonical capsids <i>do</i> represent group 1 capsids . . . . .	45
3.5	Rationalizing the need for a five-sided subunit . . . . .	46
3.6	Breaking the canonical rules . . . . .	47
3.7	Future applications . . . . .	48
3.8	Conclusion . . . . .	49
<b>4</b>	<b>Geometric properties of the virus capsid</b>	<b>51</b>
4.1	Introduction . . . . .	52
4.2	Hexamer shapes encode for capsid size . . . . .	55
4.2.1	Endo angle constraints . . . . .	59
4.2.2	Considering larger capsid sizes . . . . .	60
4.2.3	T>1 to T=1 capsid transformations . . . . .	61
4.2.4	Implications for anti-capsid therapies . . . . .	61
4.3	Endo angles and buckling transitions . . . . .	62
4.3.1	Buckling transitions versus other maturation events . . . . .	66
4.3.2	The need for auxiliary proteins . . . . .	67
4.3.3	Auxiliary proteins vs. buckling availability . . . . .	69
4.4	Concluding Remarks . . . . .	69
4.5	Materials and Methods . . . . .	70
4.5.1	Natural capsids studied . . . . .	70

4.5.2	Analysis of angles within natural capsids . . . . .	70
4.5.3	Creating capsid models . . . . .	71
4.5.4	Assaying subunit-subunit dihedral angle constraints . . . . .	72
<b>5</b>	<b>Evolution of the virus capsid</b>	<b>76</b>
5.1	The mystery of the missing capsids . . . . .	77
5.2	Revising old concepts . . . . .	79
5.3	Endo angles, classification, and history . . . . .	80
5.3.1	Endo angle propagation and termination rules . . . . .	80
5.3.2	Endo angles classify capsids . . . . .	80
5.3.3	Early attempts at size “existence theorems” . . . . .	81
5.4	Introducing Hexamer complexity ( $C^h$ ). . . . .	83
5.4.1	High $C^h$ capsids require more auxiliary control during formation. . . . .	83
5.4.2	Capsid $C^h \propto 1/\text{capsid abundance}$ . . . . .	84
5.4.3	Capsid $C^h$ is related to class $(h,k)$ not size ( $T$ ). . . . .	85
5.4.4	Evolution/natural selection vs. design . . . . .	87
5.5	Understanding the rule breakers and charting a phase diagram . . . . .	87
5.5.1	Rule breakers . . . . .	87
5.5.2	Phase diagrams . . . . .	88
5.5.3	Continuum theory and the phase diagram. . . . .	90
5.6	Further implications . . . . .	91
5.6.1	Classes, shapes and buckling . . . . .	91
5.6.2	$T$ -switching and pleomorphy . . . . .	92
5.6.3	Non-icosahedral capsids . . . . .	92
5.7	Take home . . . . .	93
	<b>Bibliography</b>	<b>93</b>

<b>A</b>	<b>Appendix: Capsid Geometry</b>	<b>103</b>
A.1	Canonical intra-pentameric dihedral angles interact at $\sim 138.19^\circ$ . . . .	103
<b>B</b>	<b>Appendix: Capsid Periodic Table</b>	<b>109</b>
B.1	Canonical vs. noncanonical capsids . . . . .	109
B.2	Defining hexamer complexity $C^h$ . . . . .	110
B.3	Counting hexamer shapes . . . . .	111
B.4	Capsids with low $C^h$ are preferred . . . . .	111
B.5	Observed capsid abundance $\propto 1/C^h$ . . . . .	113
B.6	Is there a data-collection bias? . . . . .	113
B.7	Basic definitions . . . . .	114
B.8	Obtaining endo propagation length $\phi_{h,k}$ . . . . .	115
B.9	Obtaining $C^h$ from $h, k$ and $\phi_{h,k}$ . . . . .	116
B.10	The number of hexamers $N_x$ . . . . .	119
B.11	The viruses used in calculating capsid abundance in Chapter 5 . . . . .	120

# List of Tables

3.1 Two major virus capsid groups . . . . . 30

B.1 The distribution of capsid sizes into three classes . . . . . 112

B.2 Hexamer complexity,  $C^h$  . . . . . 118

# List of Figures

1.1	Virus life cycle . . . . .	4
1.2	Functional energy landscape . . . . .	6
2.1	A range of capsids . . . . .	10
2.2	Making capsid models of various sizes . . . . .	13
2.3	Rotational symmetry . . . . .	16
2.4	Platonic solids and their cubic symmetries . . . . .	17
2.5	Quasi-rotational symmetry . . . . .	18
2.6	Chirality and mirror symmetry . . . . .	19
3.1	The tilable and the not so tilable . . . . .	32
3.2	Calculating % holes and overlaps . . . . .	33
3.3	Testing for monohedral tiling - I . . . . .	34
3.4	Testing for monohedral tiling - II . . . . .	35
3.5	A representation of the subunit tile and the capsomers . . . . .	38
3.6	Pentagonal type 5 tilings . . . . .	42
3.7	Mirror symmetries preclude 3 and 4 sided subunits . . . . .	47
4.1	Two capsid sizes . . . . .	54
4.2	Hexamer shape is specific to capsid size . . . . .	57
4.3	Simple geometry describes hexamer shape . . . . .	64

4.4	Only $T \geq 7$ capsid models appear to “buckle” . . . . .	66
4.5	Squeezing and stretching dihedral angles . . . . .	73
5.1	The inconspicuously missing capsids . . . . .	78
5.2	Endo angle termination rules . . . . .	81
5.3	The three virus capsid classes . . . . .	82
5.4	Evolutionary discrimination of spherical capsids . . . . .	84
5.5	Periodic table of spherical capsids . . . . .	86
5.6	Spherical capsid phase diagram . . . . .	89
A.1	Showing that $\phi \sim 138.19^\circ$ . . . . .	104
A.2	Hexamer angle profiles . . . . .	107
A.3	Capsid model buckling profiles . . . . .	108
B.1	Hexamer shapes available to capsids . . . . .	110
B.2	Capsids tend to prefer lower $C^h$ . . . . .	112

## ABSTRACT

# Principles of Virus Capsid Design

Ranjan V. Mannige

The Scripps Research Institute

Doctor of Philosophy

The survival of most natural viruses is dependent upon the existence of spherical capsids—shells of various sizes composed of protein subunits, which serve as a protective coat for the virus. Since capsids are employed in almost all aspects of the viral life cycle, understanding both structural and dynamical features of capsids remain imperative. In this thesis, we will employ theory into understanding such properties.

From a geometric, topological and physical perspective, we uncovered aspects of the spherical virus capsid at various levels of structure (subunit properties [1] and capsid scalability [2]), function (maturation [2] and rigidity [2, 3]), design (from first principles [1–3]) and evolution [3].

The resulting theories show that virus capsids, although famously diverse, may be unified by a mathematical framework (which culminates in a periodic table). This framework provides an opportunity to explore a large number of capsids to benefit our understanding of an integral player in the virus life cycle, while informing the field of nanotechnology of general assembly design requirements. Our hope is that these theoretical understandings may further be employed in designing antiviral therapeutics and completely artificial high-order molecular assemblies.

# Chapter 1

## Introduction

This document strives to be a *narrative* on the hunt for virus capsid design criteria, and is designed to be read and understood by anyone. I hope that it will be readable in bed without a pencil.

### 1.1 Viruses: a general introduction

Much before their visual discovery, spherical viruses (e.g., smallpox, yellow fever) had already remodelled human life, a pithy case of “what you don’t know *can* hurt you”<sup>1</sup>. But their discovery, the realization that such nefarious organisms barely even passed for “living”, shook biology as a field, and since then, viruses have regularly provided us with new surprises. For example, recently, scientists discovered that the structure of many viruses infecting all domains of life—bacteria, archaea and eukarya—are remarkably similar in shape (morphology) [5]. How are such diversely found viruses so

---

<sup>1</sup>Smallpox is believed to have redrawn history as early as 10,000 BC [4], while yellow fever resistance was one of the deciding factors in the unfortunate slave trade trades of the later years.

similar? This question, some scientists contend, lies at the center of our understanding of the beginnings of life on earth [6], and will be discussed in the coming chapters. But we are getting ahead of ourselves.

Viruses come in a variety of shapes: many regular or highly symmetric (e.g., the spherical and rod shaped viruses), some more imaginatively shaped (like the “lemon”, “cone”, and “hotdog” shapes [7]), whereas others are thoroughly messy and come in a variety of irregular awtars, only one of which may be virulent (the ebola virus is a prime example, with its characteristic “shepherd’s crook” shapes). The shapes of many of these viruses are defined by the shape of the capsid—a protective shell that shelters their genomes and facilitates virus transportation. These capsids are primarily composed of protein subunits that come together (“assemble”) spontaneously in the correct setting/environment, which is quite a feat!

Only a subset of virus capsids—the spherical ones—have so far afforded thorough structural analysis; their highly symmetric nature, for various practical reasons, has allowed for their extensive study in the last 50 years from biochemical, physical and structural perspectives [8]<sup>2</sup>. On account of this diverse available dataset, we have chosen to understand design criteria of spherical capsids, and will focus primarily on them throughout this dissertation. It is, however, possible that rules gleaned from the study of spherical capsids allow for a greater understanding of capsids of non-spherical shapes (which will be focused on later). In the coming sections, we will review the concepts that motivate these investigations.

---

<sup>2</sup>This is for two major reasons: high symmetry allows for the easy production of three-dimensional models from cryo-EM and X-ray crystallography methods (since various equivalent points in the capsid can be averaged), while the ability to pack regularly into crystals further enhances the possibility of obtaining atomic resolution x-ray crystal structures.

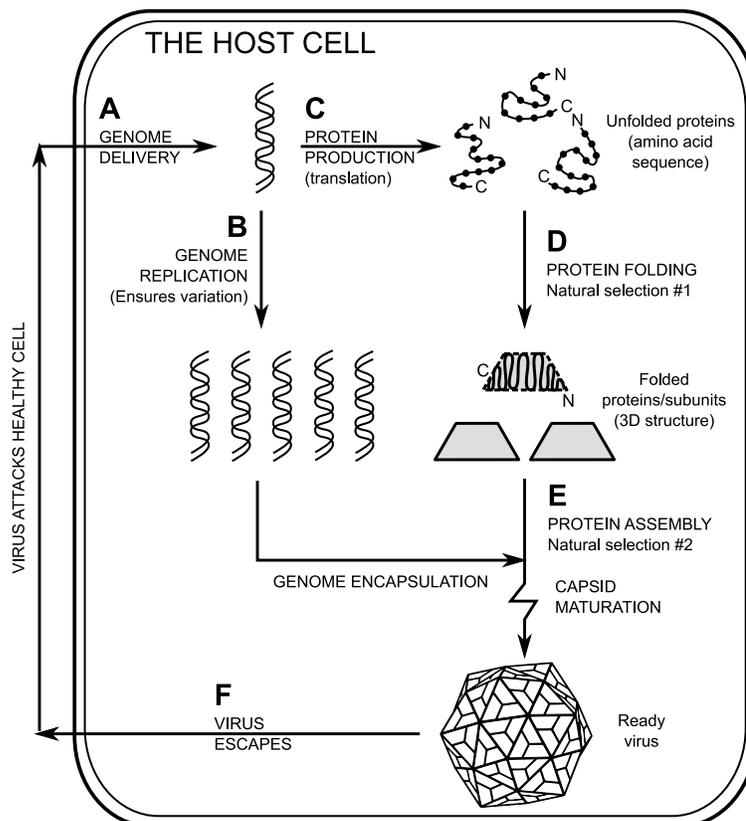
## 1.2 Specific reasons to study the capsid

I feel that the capsid's intricate beauty and symmetry are enough to elicit infinite forays into the capsid world (it should be obvious by now that I like looking at them). However, there are more practical (i.e. grantable) properties of capsids that motivate investigations into capsid structure and function:

(1) *Capsids are engineering feats.* Most engineers would be very interested to know how capsid subunits self-organize (assemble) into highly symmetric shells. This is particularly because capsids do this activity with very high fidelity (they come together to form the *correct / native* capsid structure without making many mistakes, i.e. without forming other structures such as broken capsids and what can scientifically only be termed as “goop”). That capsids exist at the nanometer size-scale makes these self assembling structures even more useful as wise teachers of the design of artificial nano-assemblies.

(2) *Capsids are useful scaffolds.* They serve as novel platforms on which interesting chemistry and immunology can be performed. Particularly, capsids are useful for these purposes because “decorating” a capsid *subunit* with specific functionalities means that that functionality increases sixty-fold in the final capsid (more on capsid structure in Section 2.1.3).

(3) *Virus life cycle is capsid centric.* The capsid is often a virus' sole armor without which the virus dies. But beyond just protection, capsids hold useful roles in almost all other aspects of the virus lifecycle, discussed in Figure 1.1. The more we know about the capsid, the more viruses we can break on purpose.



**FIGURE 1.1: Virus life cycle.** Generally, the virus lifecycle proceeds with the free virus encountering a suitable host cell and then “internalising” its genome into the cell by various means (A). Once the virus genome interacts with the cells replicative machinery, more viral genomes (B) and capsid proteins (C,D) are produced that assemble into newly formed viruses (E) that escape the cell (F) to invade other cells. Most of these activities involve the virus capsid, which underlines our need to understand these little devices.

In summary, understanding how capsids work may inform us of new ways of making, breaking, modifying and copying capsids. Specifically useful in harnessing virus capsids are the physical properties of the subunit and the final capsid (assembly), and the ability of the subunit to interchange between both free and associated forms. All these things we term as the *design criteria for virus capsids*, and we hope to understand these properties from a mathematical perspective. For that purpose, the fol-

lowing section has been added to (hopefully) dispel the notion that all biology is not amenable to mathematical treatment.

### **1.3 Math and biology can be friends!**

Historically, math and physics have been progressing hand in hand quite nicely. Unfortunately, this can not be easily said for biological systems: although math has been utilized as a tool in gathering data (e.g., the Fourier transformation is used in structure determination) math's utility in *explaining* and *predicting* biology has been limited at best<sup>3</sup>. Still, a few biological features/systems are described by “elegant” math/theory<sup>4</sup>, for example, Mendel's genetic rules. Which biological systems are intractable and which are mathematically lucid? This section provides the abstract means to look at a biological system and answer the above question. I hope that, at the end of this section, the reader will be amenable to the application of geometry and mathematics in the understanding of the structure and function of virus capsids.

#### **1.3.1 Functional energy landscape**

The mathematical accessibility to some biological systems may be assessed by evoking the *functional energy landscape* (Figure 1.2). It is a conceptual “genome space” that is sampled by similar organisms competing for survival under a functional pressure.

---

<sup>3</sup>That is ironic, since many of the major progresses in biology were made by people that would traditionally be described as physicist-types.

<sup>4</sup>I describe “elegant” as “not *too* complicated” or “void of knobs, switches, buttons, and tweaks”.

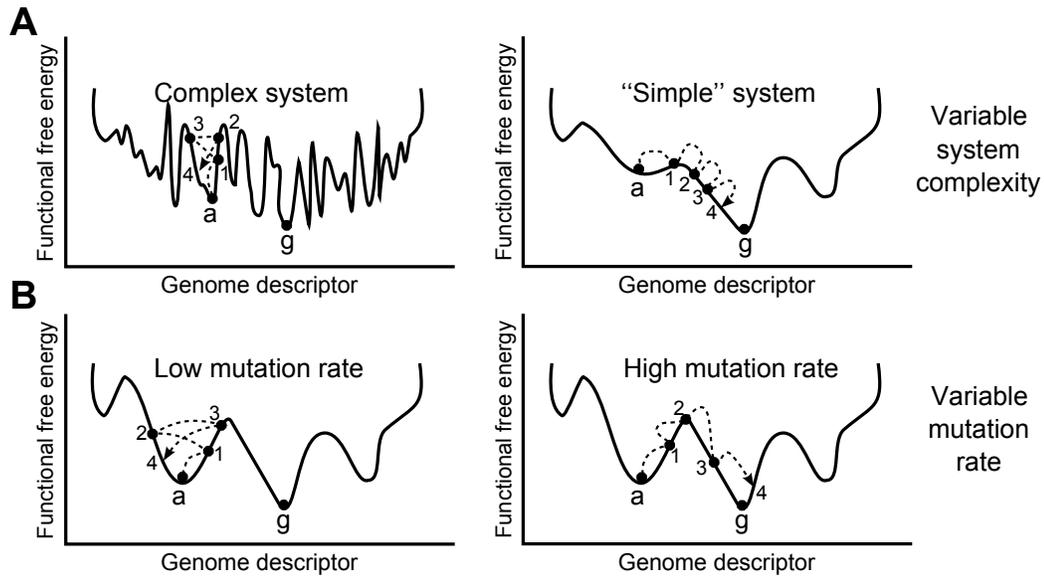


FIGURE 1.2: **The functional energy landscape.** The nature of the functional energy landscape's complexity (A) and the average rate of mutation for the genome (B) affect the chances of a genome in encountering the best design possible (the global minimum, **g**) from its current situation **a**. Viruses are expected to have both low complexity and high mutability, which is what, we propose, make their capsids amenable to description by mathematics.

The X-axis is a hypothetical unit that describes the genome/gene of the organism that is involved in providing the function in consideration<sup>5</sup>. The Y-axis describes the functional free energy. The lower this value, the more efficient the organism is at that function, i.e., genomes will tend to collect at the landscape minima<sup>6</sup>.

### 1.3.2 Making the mathematical connection

It is often true that the easiest way to do something well (something "not too complex") is often mathematically elegant and predictable from first principles. For example, if we fold a paper onto itself, the crease that arises is predicted to be a line<sup>7</sup>, which is also

<sup>5</sup>It is obvious that organisms of similar genomes will describe similar functional efficiencies and will therefore probably belong to the same well in Figure 1.2.

<sup>6</sup>Incidentally, the competing models of evolution—cladogenesis by *punctuated equilibrium* and *phyletic gradualism*—are supported by a smooth and a rough landscape respectively.

<sup>7</sup>This is a salient proof that arises from Euclidean geometry.

the case in reality<sup>8</sup>. Although this sounds obvious, it is important to note that some mathematical predictions, although not intuitive, are still simple and elegant (e.g.,  $e = mc^2$ ). Let's push this line of thought with the conjecture that *the global minimum to a functional energy landscape describes genomes whose products are mathematically elegant*.

The ability of an organism to *arrive* at the most efficient design (i.e., the global minimum, and hence the mathematically describable solution) will require the following properties: (1) the organism's genome must be mutable, without which, sampling of the functional landscape is impossible, (2) higher mutation rates will result in higher genome sampling thereby increasing the organisms chances in encountering the global minimum, and (3) optimally, the landscape must have only a few deep minima, i.e. the "system complexity" must be kept low<sup>9</sup>. These are all acceptable criteria to anyone who uses Monte Carlo simulation methods to search for a global energy minimum.

***Proposal:*** We propose that the chances are high that the structure and function of virus capsids<sup>10</sup> may be mathematically accessible. This claim emerges from the knowledge that viruses display (1) high genome mutation rate, (2) low system complexity (compared to other biological phenomena) and (3) high population doubling rate. Aware of the virus' relative simplicity, this thesis asks the following question throughout the chapters: can efficient virus designs be explained by simple mathematical principles? We direct this question first to the subunit shape itself (Chapter 3),

---

<sup>8</sup>However, the crease patterns of an unfolded origami crane is not so easily predicted as it is more complicated (but the lines are still straight!).

<sup>9</sup>A complex system has more degrees of freedom and therefore will tend to display a larger number of deep local minima.

<sup>10</sup>That are present on account of a rather strong evolutionary pressure based on the need to protect and encapsulate the viral genome.

then to the capsid assembly (Chapter 4) and finally to the natural selection of capsids (Chapter 5).

## 1.4 Limitations of math

The difference must be made between features of a capsid that are general and those that are host-specific. General properties are those that are not intimately associated with the host. For example, while the capsid subunit performs the *general* mechanical function of assembly and disassembly, certain features available *on* the capsid surface may be directly involved with interacting with the host to facilitate genome entry (Figure 1.1A) or escape (Figure 1.1F). Such aspects of the capsid that interact with the host may be mathematically intractable since the system size becomes larger (as the host would have to be “included in the equation”). This is reflected in the diverse methods of viral genome internalization (Figure 1.1A) that are host specific<sup>11</sup>.

It is this limitation of mathematics/theory that underlines the dire need for close interactions between experimentalist (who can explain such things as capsid-host interactions) and theorists (who can explain general design criteria and general vulnerabilities of capsids). Both are interesting and crucial to the fight against viruses!

Before we head on to the (hopefully) fun theoretical questions, the next chapter will discuss topics that bring everyone up to speed.

---

<sup>11</sup>Methods used by viruses include internalization (1) by passive means (where the virus enters the cell through a scratch/wound, e.g. plant viruses such as SMV and TBSV), (2) via the endosomal pathway (many animal viruses including the polio and the common cold or rheovirus), (3) by genome injection (tailed bacterial viruses such as HK97), and (4) by a combination of methods (e.g., the herpesvirus, which passes through the cell wall by method #2 and then finally injects its genome into the nucleus in a manner reminiscent of method #3).

## Chapter 2

# Introduction II: Selected Topics

This is possibly the most boring chapter in this document. Why? Because it is the one chapter that will be referred to repeatedly in the remainder of the book (so, its also quite important). Topics include symmetry, the history of early virus capsid theories, the famed triangulation number ( $T$ ), quasi-equivalence, biology's role in nanotechnology, *et cetera*. In case of extreme boredom/restlessness, swiftly proceed to Page 24.

### 2.1 Structural virology: then and now

In the early 1950s, very little was known regarding spherical capsid structure<sup>1</sup>. Then, in the span of six years, our understanding of spherical virus capsid structure grew by leaps. The two important milestones lay in realizing that (1) spherical capsids are made up of subunits and possess icosahedral symmetry (Section 2.1.1) and (2)

---

<sup>1</sup>Mostly, we knew that they were made of protein.

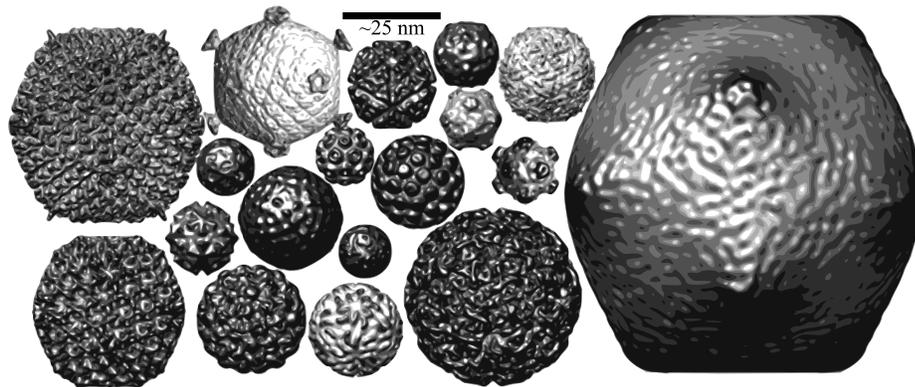


FIGURE 2.1: **A range of capsids** (shown roughly to scale) collected from the VIPER EMdb website [8].

capsids are scalable (Chapter 2.1.3). These two ideas have allowed us to understand and structurally characterize all spherical capsids that we see today (Figure 2.1).

### 2.1.1 Milestone I: Equivalence and symmetry

Even before protein subunits could clearly be seen within the confines of a capsid, an interesting and important puzzle had presented itself: it was becoming clear that the the size of any capsid was *much* larger than the largest protein that the enclosed viral genome could express/produce (Remember, the genome enclosed within the capsid must possess the blueprints for the capsid itself). How could this be explained?

Crick and Watson, after their Double Helix success, reasoned that one could form such a capsid only if viruses have figured out a way to arrange multiple copies of a smaller protein (a “sub”-unit) into the form of a shell<sup>2</sup>. This is where *symmetry* and *order* comes into the picture (discussed in Section 2.2). Based on crystallographic evidence [10], Crick and Watson had proposed that the capsid would have to assume

---

<sup>2</sup>They they noted that “The question we must now ask is whether the protein shell of the spherical viruses is likewise constructed by a regular aggregation of one type of small molecule and if so, how this is done.”(Pg. 474 in Ref. [9])

a high order symmetry group (Section 2.2.2). In doing so, large copies of the same subunit would possess identical or *equivalent* positions within the capsid (hence the idea of *equivalence* between the subunits) The proposed symmetries were the ones displayed by platonic solids (Figure 2.4) [9], of which the icosahedral symmetry is the highest in order.

To this day, almost all spherical virus capsids possess icosahedral symmetry, which must have been quite pleasing to Crick and Watson (and is a triumph of theory-driven science!). However, new methods (such as negative staining EM) soon showed that the number of subunits per capsid were in slight disagreement to the Crick-Watson proposal. In stead of an icosahedral structure with sixty equivalent subunits, spherical capsids, albeit icosahedrally symmetric, were found to be composed of *multiples* of sixty subunits! How could capsid subunits perform such feats of scalability?<sup>3</sup>

### **2.1.2 Milestone II: Scalable capsids (a confluence of disciplines)**

Horne and Wildy, in 1961 [11], had noticed that subunits within capsids appeared to exist in clusters of five and six subunits<sup>4</sup>. They noticed that these structures looked quite similar (in an abstract sense) to the structures that Buckminster Fuller used to make [11]. In those days, Buckminster was well into his hobby of putting hexagons and pentagons (*gons* not *mers*) together to create rather robust spherical structures.

Although Horne and Wildy noticed Buckminster's architecture [11], it was Caspar and

---

<sup>3</sup>In fairness, CW did not ignore the possibility of capsids with multiples of 60 subunits, they just didn't provide a method to understand such configurations. This is reflected in their line: "... further points must be made to prevent misunderstanding. ... [the subunit need not be] a single protein molecule in the chemist's sense of a unit joined together by chemical bonds. Several different protein molecules may aggregate to form the asymmetric unit" [9].

<sup>4</sup>Later, these clusters were called pentamers and hexamers respectively.

Klug who actually went back to Buckminster’s personal notes, which resulted in the seminal paper that introduced the idea of the triangulation number, *quasiequivalence*, and the scalable capsid (Section 2.1.3) [12]. From that point, Structural Virology was conceived in the form that persists even today.

**A meander.** Interestingly, Caspar and Klug’s method for creating scalable capsids had been introduced a whole thirty years before by the geometer Goldberg [13], who had shown a way to create an unlimited range of polyhedra of any platonic symmetry with the help of two integers ( $h,k$ ; described in Section 2.1.3). If only biologists had read this gem, the observation that viruses existed in various sizes (i.e. that they are “scalable”) would have been a lot less mysterious. Alas, the records of the Japan-based Tōhoku Math Journal were lost in US bombing raids of WW-II and only post 1949 volumes are available online. Could that possibly be the reason for Goldberg’s lack of recognition by the virus community?

### 2.1.3 Making “scalable” capsids

In stead of simply explaining the Caspar-Klug (and Goldberg) method for creating capsids of various sizes, I thought of recounting Jack Johnson’s enthusiastic hands-on “lab” section to the Structural Biology class in Scripps (which is an anecdotal version of the method anyway). We started the class with large sheets of yellow chartpaper on which a hexagonal lattice was printed (Figure 2.2A). We were then allowed, in groups, to convert this boring piece of chartpaper into exciting capsid models of selected size (Figure 2.2C; size was selected by choosing two integers “ $h$ ” and “ $k$ ” that were first introduced by Goldberg, and hence will be referred to as Goldberg integers).

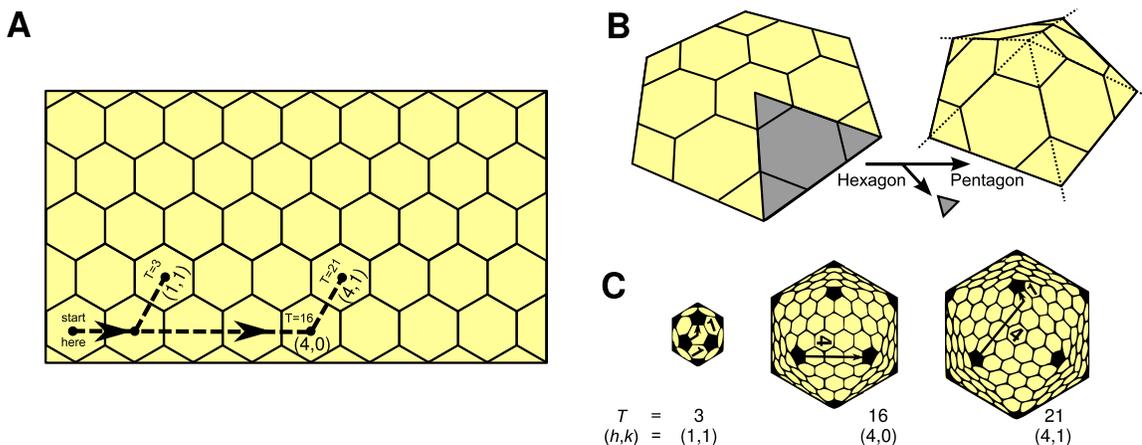


FIGURE 2.2: **Making capsid models of various sizes described by  $(h,k)$  pairs.** The general idea is that all capsids consist of 12 pentamers (darkened in **C**) and a variable number of hexamers. One may go about creating paper models of such capsids by starting out with only a sheet of hexagons (**A**), where hexagons represent hexamers, and then selectively converting specific hexamers into pentamers (**B**).

This transformation was done by walking along the chart paper from doomed hexagon to doomed hexagon<sup>5</sup> (they were doomed because they would be soon converted to pentamers, shown in the final capsid models in Figure 2.2C as darkened pentagons). These hexagons were converted into pentagons by excising  $1/6^{\text{th}}$  of the selected hexagon (and its neighboring paper) and gluing the unpaired edges (Figure 2.2B). Make 12 such pentagons, join the unpaired edges, and you have a three dimensional model of a capsid, where pentagons and hexagons represent pentamers and hexamers respectively.

What was important about this activity is that it allows one to see that capsid size is completely defined by the  $(h,k)$  pair. Also, we arrive at an understanding of how real capsids can comprise exclusively from pentamers and hexamers.

<sup>5</sup>Based on our selected  $(h,k)$  pair, we would have to “walk”  $h$  hexagons in one arbitrary axis on the paper and  $k$  hexagons in an axis that is placed at a  $(2\pi)/6$  radian angle to the first axis.

### 2.1.4 Quasi-equivalence in virus capsids

Although creating scalable polyhedra (by means of two integers,  $h$  and  $k$ ) was not a new idea at Caspar and Klug's time, their idea of *equivalence* was original and still is hailed as a breakthrough in structural biology. They stated that one could produce both hexamers and pentamers from the same subunit, provided that they exist in two distinct but similar (or *quasiequivalent*) environments<sup>6</sup>. When the subunit's global structure itself is not drastically changed, this quasi-equivalence is manifested in the subunits ability to interact with the same partner in more than one way, making that subunit-subunit interface a quasi-equivalent interface. One could use Figure 2.2B to imagine how this is done in real viruses: instead of a  $1/6^{\text{th}}$  wedge being excised, we would only have to jettison one protein subunit from the hexamer and let the unpaired interfaces interact to form a pentamer (shown in Figure 5.2B).

Today, quasi-equivalence is a phenomenon that is believed to be ubiquitous to biological systems<sup>7</sup> and is seen in processes as diverse as icosahedral virus capsid assemblies [14, 15] and propagation of signal cascades [16, 17].

### 2.1.5 Triangulation ( $T$ ) number

Although  $h$  and  $k$  are useful in understanding capsid size and arrangements of pentamers and hexamers, its not always convenient to deal with two numbers as a descriptor. Conveniently, Caspar and Klug presented a handy number, the *triangulation*

---

<sup>6</sup>Caspar and Klug stated that "The basic assumption is that [the] shell is held together by the same type of bonds throughout, but that these bonds may be deformed in slightly different ways in the different non-symmetry related environments."(Page 10 of [12])

<sup>7</sup>As often protein molecules manifest multiple quasiequivalent configurations that interact with their partners in multiple ways.

number,

$$T = h^2 + hk + k^2. \quad (2.1)$$

$T$  in Equation 2.1 is useful because it easily describes the number of subunits ( $60T$ ) and hexamers ( $10[T - 1]$ ) in the capsid and the number of subunits in the asymmetric unit (which is  $T$  itself!). Today, the triangulation number is an unalienable and expected descriptor of any spherical capsid.

### 2.1.6 Mini-summary

Spherical capsids of all observed sizes may be obtained from a grouping of twelve pentamers (symmetric clusters of five subunits) separated by a variable number of hexamers (clusters of six subunits) [11, 12] represented in Figure 2.2C.

Capsid size may be characterized by two integers,  $h$  and  $k$  (first discussed by Goldberg [13] and then by Caspar and Klug in the context of capsids [12]).

A capsid of triangulation number  $T$  is comprised of  $60T$  subunits, or 12 pentamers and  $10(T - 1)$  hexamers, i.e.,  $T$  is a quantitative metric for capsid size (still, we now show in Chapter 5 that  $h$  and  $k$ , and not  $T$ , are more useful in understanding capsid classification and evolution).

## 2.2 Symmetry

### 2.2.1 Rotational symmetry

Rotational symmetry is prevalently visible in macromolecules of all types. This is especially true for viruses, where capsid subunits are known to assemble into (rotationally)

symmetric clusters of two, three, five or six subunits (called “capsomers”) that then come together to form the final capsid. Lets look a little more at this kind of symmetry.

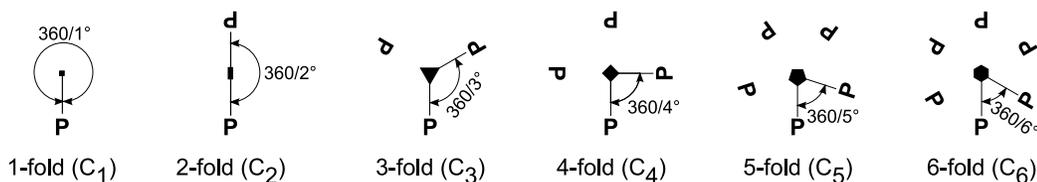


FIGURE 2.3: **Rotational symmetry.** Sets of ‘P’s that are related by  $C_1$  through  $C_6$  symmetry. The central symbol denotes the point through which the symmetry axis falls perpendicularly through the paper. An operation around a  $C_n$  axis is a rotation that is made in multiples of  $360/n^\circ$  or  $2\pi/n$  radians.

Rotational symmetry (denoted as  $C_n$ , where  $n$  is an integer; see Figure 2.3) deals with mapping one part of an object or cluster to another with the help of a “pole” or symmetry axis. A rotational symmetry *operation* occurs when one, as the observer, holds the pole with one hand and moves round the pole from one frame of view to another such that, after the operation, nothing in the system appears to have changed<sup>8</sup>. So, a symmetry operation is one where, after the operation is made, the observer is left with a strong feeling that he/she has seen “exactly the same thing before”<sup>9</sup>.

## 2.2.2 Platonic/High order structures

The **order** of any symmetry is the distinct number of times the asymmetric unit (“P” in Figure 2.3) is exhaustively repeated by the available symmetry operations. So, if no other symmetry elements exist, the order of a  $C_n$  symmetric structure is  $n$ , and that of

<sup>8</sup>Now, imagine the number of  $C_1$  rotational symmetry operations that are happening as we speak in Las Vegas! That could be your homework.

<sup>9</sup>Experts say that *déjà vouz* could be a result of such symmetry operations in life.

a mirror symmetry is two<sup>10</sup>. The order of a system gets harder to predict when we go to larger groupings of symmetry elements.

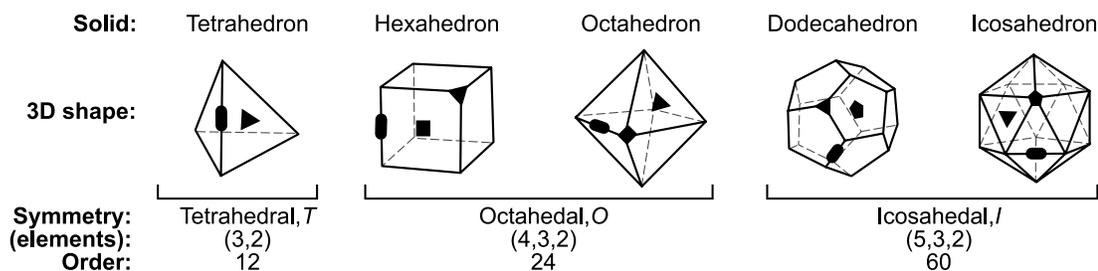


FIGURE 2.4: Platonic solids and their cubic symmetries. The word “cubic” is used because the  $C_3$  element is the largest common symmetry element seen in all three symmetry types.

*Platonic solids and icosahedral symmetry.* Interesting things happen when one groups a more complicated bunch of symmetry elements together in space, especially if those symmetry elements intersect at a point<sup>11</sup>. For example, if done wisely, we can increase the order of a system by cleverly grouping rotational symmetry elements together to obtain the following “spherical” structures of high order (Figure 2.4): the tetrahedron (possessing specifically arranged 3 and 2-fold rotational symmetry elements that define *tetrahedral symmetry*,  $T$ <sup>12</sup>), cube and octahedron (both possessing identically placed 4,3, and 2-fold rotational symmetry elements that define *octahedral symmetry*,  $O$ ), and the icosahedron and the dodecahedron with 5,3, and 2-fold rotational symmetry elements (*icosahedral symmetry*,  $I$ <sup>13</sup>).

<sup>10</sup>Order can also be thought of as the number of instances of the asymmetric unit in the symmetric system (the number of symmetry-relatable “P”s in a cluster in Figure 2.3, for example)

<sup>11</sup>We have inadvertently described the *point group* that is an integral component in chemistry and structural biology.

<sup>12</sup>Not to be mistaken for *triangulation number*,  $T$ ; Section 2.1.5

<sup>13</sup>It should now be obvious that an object displaying *icosahedral symmetry* (e.g. the dodecahedron) need not *look* like an icosahedron, and an object displaying *octahedral symmetry* (e.g. the cube) need not *look* like an octahedron. This is an important concept that lends an understanding into how virus capsids that are predominantly icosahedral in symmetry can take on many structural forms.

The  $T$ ,  $O$  and  $I$  symmetries possess orders 12, 24 and 60, respectively. So, if we tried, we could stuff 12, 24 or 60 subunits in identical (or *equivalent*) environments into a shell by placing them in  $T$ ,  $O$  or  $I$  symmetric arrangements, respectively. This is a very important concept that was used by Crick and Watson [9], who reasoned that spherical capsids may be forming from identical subunits that assemble into, in fact,  $T$ ,  $O$  or  $I$  symmetric forms (Section 2.1.1).

### 2.2.3 Quasi-rotational symmetry

Most six-subunit clusters (“hexamers”) in virus capsids *appear* to possess  $C_6$  symmetry even though they don’t (remember, an icosahedrally symmetric structure does not allow  $C_6$  symmetry elements). We say that these subunits possess *quasisixfold*-symmetry. This happens on account of *small* changes that exist between subunits within the cluster that destroys the sixfold symmetry. Examples of this happening given our system of “P”s is shown in Figure 2.5B-E, where the slight differences in the color of the “P”s result in the  $C_3$ ,  $C_2$  and  $C_1$  clusters that *appear* to possess six-fold ( $C_6$ ).

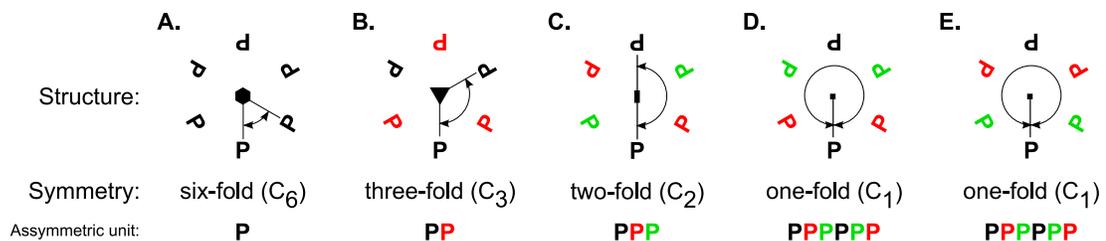


FIGURE 2.5: **Quasi-rotational symmetry.** Although all these clusters appear to have six-fold rotational symmetry ( $C_6$ ), only (A) does, while the rest don't on account of differences in color. (D) and (E) are  $C_2$  transformations of each other and are not identical (or superimposable), which emphasizes that the structure does not even have  $C_2$  symmetry, let alone  $C_6$  symmetry.

In capsid hexamers, the  $C_6$  symmetry is broken not by the existence of different “colored” subunits, but often by the different *angles* at which those subunits interact (the hexamers need not all interact with each other at planar angles. they have 3D shape! See Figure 4.3C). This feature of hexamers will turn out to be very important in understanding how capsids size (Section 2.1.3) may be controlled (discussed in Chapter 4).

#### **2.2.4 Chiral subunit clusters can not possess mirror symmetries**

Another concept that will help us in a later chapter (Chapter 3) is that virus capsids (and other biological structures) can not possess mirror symmetries on account of their chiral nature<sup>14</sup>.

The asymmetric unit ‘**P**’ in Figure 2.3 was used for a specific reason: it is chiral (it can possess no mirror symmetry elements), just like proteins<sup>15</sup>. In Figure 2.6, we use a chiral (‘**P**’) and achiral (‘**A**’) asymmetric unit to emphasize that mirror symmetry elements (dashed lines) may exist in the symmetric cluster only if the asymmetric unit itself allows for mirror symmetries (is achiral), i.e., protein clusters can not possess mirror symmetry elements (or mirror planes).

The exclusion of mirror symmetry from protein clusters will manifest itself in interesting ways when we look at the design criteria for the subunit shape (Chapter 3).

---

<sup>14</sup>Being chiral means that the object and the image of that object once seen through a mirror are distinct, which means that such objects can not host mirror symmetry elements.

<sup>15</sup>Proteins are fundamentally chiral due to their chiral polypeptide backbones.

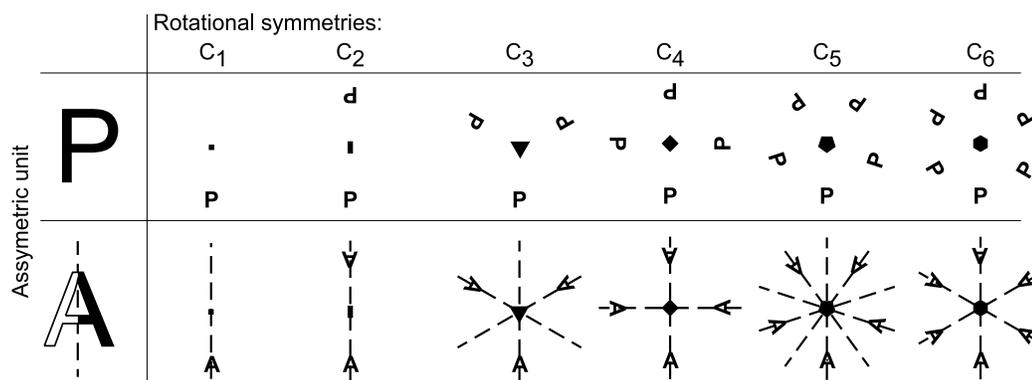


FIGURE 2.6: Structures may include mirror symmetries only if the *asymmetric unit* contains a mirror symmetry element (dashed lines).

## 2.3 Introduction to nanotechnology

Science, by many, has been described as a continual search in two directions—inward and outward. Although this sentence is ambiguous and has many meanings, all the interpretations of this phrase, I believe, are most evident today. There are scientists looking inwards at the mysteries of the human mind and others looking at society as a collection of minds. There are machines currently probing both the vastness of space and the equally “vast” regions of the sub-nuclear universe. Although much of the research done in these bleeding-edges of science are intangible to the public, one field—Nanotechnology—is not only tangible, but steadily gaining fame as the next big revolution that will affect almost every aspect of personal life and industry.

**Definition.** Nanotechnology definitions come aplenty, and most often reflect the definer’s school of thought. Often, when asked about nanotechnology, chemists think of buckey balls and nanotubes, electrical engineers think of nano-lithography and silicon chips, and biologists think of proteins. Really, it is anything to do with technology

at the nanometer scale (one billionth of a meter)<sup>16</sup>. Controlling matter at that scale would mean atomic control the likes of which, I feel, parallel the dreams of alchemists.

Often, nanotechnology is thought of as the control of matter from two independent approaches: (1) top down and (2) bottom up. The former deals, essentially, with chiseling, etching, or lithographing bulk materials into displaying nano/micrometer-scale features. That is how computer chips are made. The bottom-up method is more interesting to me because, instead of employing big costly machines to do all the building work (e.g. those housed in lithography clean rooms), the molecular raw materials itself (subunits) are allowed to toil away by assembling together into “machines” that perform the desired functions. Just like virus capsids.

## **2.4 Nanotechnology, history and viruses**

Modern nanotechnology’s conception emerged from a collection of papers released in the 1960’s of which Feynman’s is most well-noted. In the 1960’s and 70’s, due to shortcomings in tools and technology, instead of gaining momentum in the scientific community, Feynman’s nanotechnology seeped into areas of art and the now famous sci-fi literature that painted futuristic worlds of nano-bots whizzing through the mind, body, factory, and planet, performing feats which would be looked upon simply as miracles.

Today, molecular biophysics–protein science in particular–is being recognized by many as an excellent springboard to strengthen nanotechnology’s hopeful ascent [18].

The atom-level enzymatic techniques and mechanics that nature has been using and

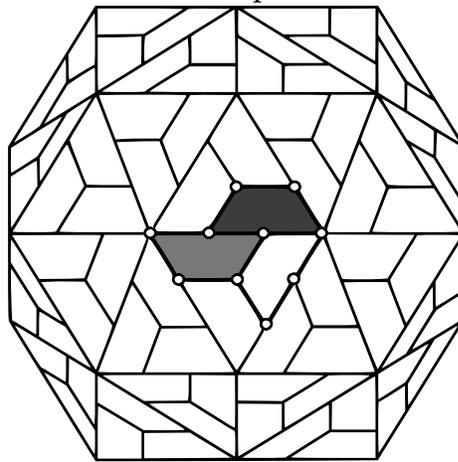
---

<sup>16</sup>For a perspective on its tinyness, if you were a nanometer tall space traveler, a cold virus would look like a big elephant, your football would be a hydrogen atom, and you could colonize a human eyeball with all your friends, for it would be the size of three of your earths.

refining for billions of years will serve as important lessons in our final aims of designing and building those nano-industries that were envisaged by Feinman, Drexler and the slew of sci-fi writers. Indeed, the Protein Data Bank (PDB) today has is a library of more than 58,588 “nano-bots”, or proteins, that will certainly be able to teach us volumes about practical, reproducible and functional nanotechnology.

Although most of the entries in the PDB are low order assemblies containing one or a few subunits, Virus capsids that assemble from hundreds of subunits are among those rare, well honed high-order systems that allow for investigation into “nano science”. These Machines are thorough in their instructions: little subunits interact with each other to form a final structure with high fidelity. How? What are the design constraints? Those questions will be explored soon.

Next chapter ...



Euler is called upon

Models get attention

Subunits succumb to scrutiny

A common thread is unraveled

## Chapter 3

# Simplified Capsid Models Dictate Capsid Subunit Properties

Finally! The introductions are over and we can discuss real research. This chapter will focus on the building block of the capsid: the virus capsid subunit. What properties must the protein have in order to perform as a good capsid subunit? What are the subunit's design criteria? We will use mathematical models to infer the answers to these questions.

First, we must ask whether viruses can be described by simplified (geometric) models. We will show that virus capsids, and a lot of them at that, *can* be represented by simplistic models called *monohedral tilings*. We then will use graph theory and topology to arrive at simple predictions on the shape and nature of the virus capsid subunit. Happily, the solution converged on a subunit design (or shape or tiling) that is commonly seen within virus capsids found in all domains of life whose hosts range from

bacteria, to plants, to humans (thats one prolific subunit shape!). This suggests that mathematical constraints may exhibit dominant roles in the natural design of biological assemblies whose effects appear to easily transgress species boundaries.

The chapter is modified from the following journal publication: Mannige,R.V. and Brooks III,C.L. (2008) Tisible nature of virus capsids and the role of topological constraints in natural capsid design. *Physical Review E*, 77(54):051902. The paper was featured by the American Physical Society in the article “Piecing a Virus Together”, *Physical Review Focus*, 30 May 2008, volume 21, story 18; <http://focus.aps.org/story/v21/st18>

►► The first half of this chapter will focus on choosing the right tools (models) to investigate capsid design.

### 3.1 Part I. Modeling the spherical virus capsid

Laboratory studies of capsid structure and function are very important and can never be replaced by theory and mind experiments. However, the value of theoretical research is in supplying *predictions* into currently unexplored phenomena, *explanations* to puzzling experimental results, and *insights* into experiments that provide ambiguous answers. Much of these functions are performed with the help of a “model”: an important tool in the theorists’ belt that is essentially a facsimile of the system under scrutiny<sup>1</sup>. Models are important tools because they are expected to mimic the behavior of the system (and so can be used for predictions). So, it is inevitable that a thesis based on applying theory to capsids would require a section that asks a rather scary

---

<sup>1</sup>Examples of a “system” are the human brain, the stock market, a building, a capsid, etc.

question: are our models really representing the real virus capsids?<sup>2</sup> Lets further explore the concept of the model.

### 3.1.1 The model

#### 3.1.1.1 Approximations are inevitable

One easy generalization about theorists is that most of us are fond of approximations<sup>3</sup>. Not because approximations are cool or endearing, its because approximations are often all we have to work with! This would not be a problem in a world with unlimited resources and perfect recording devices, where we would use quantum mechanics to describe and explain everything<sup>4</sup>. In the real world, there is a slight problem: the accurate treatment of any system larger than a helium atom would take too much time to reach *any* solution, which means that one must be a little clever and make the system smaller via reasonable approximations (even a high resolution X-ray crystal structure model downloaded from the PDB, a marvel of sub-nanometer accuracy, is an approximation that assumes that an atom is a particle instead of a more complicated *system* of subatomic particles).

This requirement of approximation becomes even more evident when we consider simulating even the smallest virus capsids. Every virus capsid must be made up of at least 60 protein subunits<sup>5</sup>, each containing thousands of atoms themselves, making the analysis of structural and dynamical properties of these systems in *all-atom* form

---

<sup>2</sup>This question is only scary if its answer is negative.

<sup>3</sup>And so, the notion of the approximating theorist is perpetuated, mostly.

<sup>4</sup>In this universe, we would only have to describe the system as a bunch of waves, feed it into a computer that would then provide the answer to our question by solving the Schrödingers equation.

<sup>5</sup>This will be true for any biomolecular assembly that is icosahedrally symmetric (see Section 2.1.5 and Crick and Watson's symmetry discussions in Ref. [9]).

computationally very difficult (as the system size is very large<sup>6</sup>). For example, state of the art numerical simulations of a  $T = 1$  all-atom capsid [19] took 1.1 ns per day on 48 processors running in parallel! Although a testament to parallel processivity, the simulation was also a testament to the current inadequacies in all-atom capsid simulation, 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 (the clustering of atoms into “pseudoatoms”) reduces the system size drastically [20] but still falls short at the microsecond ( $\mu s$ ) time scale (also, these systems pose the new problem of correct parameterization of the coarse-grain force field). These complications have motivated a plethora of theoretical attempts aimed at understanding virus capsids using necessary and simplifying geometric assumptions/approximations regarding the nature of capsid assemblies [21–33]. Lets briefly discuss these models.

### 3.1.1.2 A plethora of models

Previous simplistic models were used in qualitatively *explaining* (and not predicting) specific virus capsid *phenomena*. For example, disks on a sphere were ingeniously used in explaining the emergence of icosahedral symmetry in capsids [21, 22]<sup>7</sup>. Explana-

---

<sup>6</sup>In almost any simulation method, at every “timestep” the system is paused and all the forces/energies in the system are calculated between neighboring atoms; the larger the system, the more amount of time the CPU will take to calculate the forces/energies, and the slower the propagation time of the simulation. So, you get a smaller “simulated time” per “simulation time” value.

<sup>7</sup>Similarly, simple van der Waals spheres [23], Stockmayer fluids [24], trapezoidal subunit building blocks [25, 26], tiles [27–30], and simple bonding units [31–33] 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.

tions are interesting, but they do not explain more than why we see what we see.

Predictions that can be made are too abstract.

*In this thesis, we are interested in providing not only abstract phenomenological understandings but also accurate predictions on capsid properties.*

This means that our model will have to pass the stringent test of “back-and-forth-ability” or “transferability”, where we can make seamless jumps from the simplified theoretical model to the high-resolution (all atom) model, and back, without losing crucial information regarding the topics of interest (discussed below). Such a simplified capsid model that represents a broad array of properties of many natural capsids is yet to be described. The next subsection takes a crack at that fabled model.

A good model:

- ✓ Represents the real system well.
- ✓ Is simple enough to be handled computationally (or even better: by pen and paper!).

### 3.1.1.3 Introducing the canonical capsid

Since we are interested in the manner in which the capsid is described and explained by geometry and related math, we proceeded to model the capsid as polyhedra whose individual faces represent one subunit each. We also took the liberty to assume that all subunits within the capsid are similarly shaped (i.e. the polyhedron model is a *bound monohedral tiling* which is formed from the tiling of one shape/face repeated throughout the polyhedron<sup>8</sup>). This model we call the *canonical capsid*.

---

<sup>8</sup>An example of a monohedral bound tiling is the cube that is formed from one shape (a square) repeated six times over to form a bound surface or a polyhedron.

★ *Canonical capsids*, if they exist, are those monohedral tilings that represent many natural capsids well.

The reason that we selected to model the capsid as *monohedral* are multifaceted (like our models!) but most importantly, it simplifies the problem<sup>9</sup>. Although the definition of a monohedral tiling is sufficiently defined by stating that it is a polyhedron whose faces are identical in shape, a more rigorous definition of our model is provided in the next paragraph. Please feel free to skip it.

**Denition of a monohedral tiling.** 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 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 [34]). 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 [35].

---

<sup>9</sup>Imagine asking the question: Imagine a virus model, a polyhedron, with each face wildly different in shape to the other. That’s a doosy.

★ *Canonical capsids* may be used in two contexts: they may represent either the theoretical *models* or the *natural capsids* that are representable by these models.

►► The next section describes the materials and methods used in assaying the extent to which capsids seen in nature may be represented by monohedral tilings. If inclined to skip this section, please go to Section 3.1.3 for the meat of the discussion.

### 3.1.2 Methods - I (Testing the existence of the canonical capsid)

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	71	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

TABLE 3.1: **The list of virus families** from which capsids (of triangulation numbers  $T$  and detailed in Section 3.1.2.1) were used to test for monohedral tilability (defined in Section 3.1.1.3). The families are grouped into those containing capsids that *are* and *are not* representable by monohedral tilings (group 1 and 2, respectively). “Monohedral tilability” is indicated by low values for: *a.* the extent of average subunit-subunit overlap<sub>M</sub> percent (% O), *b.* average percent holes in the capsid (% H) and, *c.* average subunit variance in the capsid in Angstrom units (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, V), and (2) capsids that cannot be represented as monohedral tilings, i.e., capsids that possess holes, gross overlaps and subunit variability.

### 3.1.2.1 Virus capsids analyzed

For our analyses (described in the next section), we used all capsids present in the VIPERdb virus capsid repository (as of April 2007) [8] comprised of chemically identical subunits with triangulation numbers greater than one. Each of the following capsids (described as PDBIDs) 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, 1mvp, 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.)

Tests for monohedral tilability:

- ✓ *Tilability* is indicated by low subunit-subunit overlaps and few holes in the shell.
- ✓ *Monohedrality* is indicated by low subunit-subunit variations in structure.

### 3.1.2.2 Test for capsid “tilability”

To see if a polyhedron can represent a virus capsid, we really are asking whether the third dimension to the virus capsid shell—the thickness—can be done away with by projecting the capsid shell onto a bound surface (e.g., the surface defined by the average radius of the capsid). This would be hard/impossible if the capsid displays a large number of holes and subunit-subunit overlaps (see Figure 3.1).

For that reason, we need to characterize the extent to which virus capsids display *holes* and *subunit-subunit overlaps*. Although this may be qualitatively done visually,

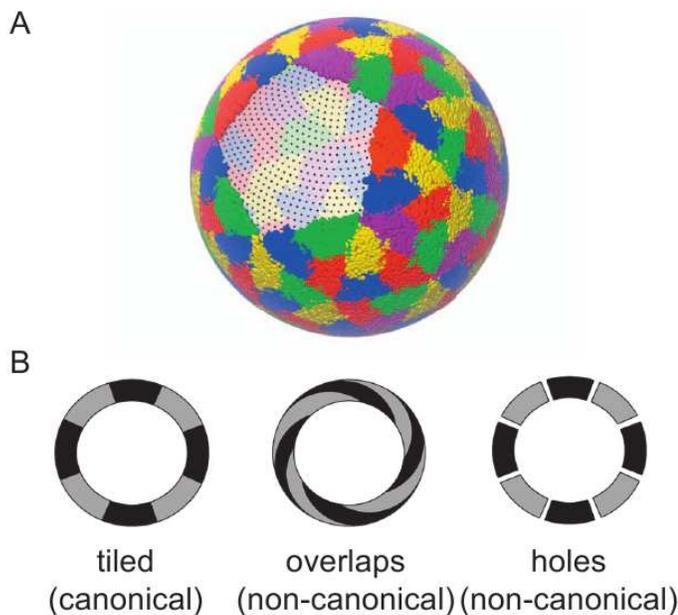
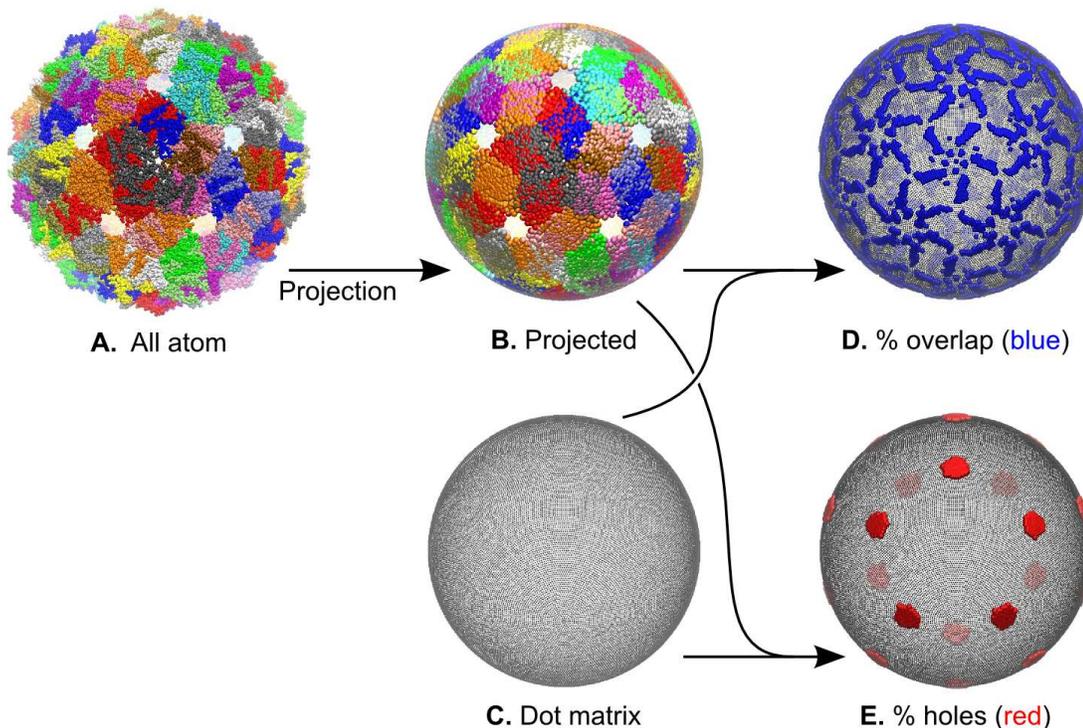


FIGURE 3.1: **(A)** The three-dimensional all-atom model is projected onto a sphere of average shell radius (colored to represent individual subunits) upon which a net, or dot matrix, is cast (visible in the subunit cleared area). These dots are used in calculating the areawise percentage of holes and 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).

we chose to develop a simple metric for quantitative characterization (Figure 3.2). 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 (Figure 3.2B). We then cast a net of dots—or a “dot matrix”—onto the shell (shown separately as Figure 3.2C) and calculated the percentage of dots that were present within holes (Figure 3.2D) and subunit-subunit overlaps (Figure 3.2E)<sup>10</sup>. These percentages were obtained for each virus capsid listed in Section 3.1.2.1.

<sup>10</sup>Each “dot” in the matrix is  $\sim 1\text{\AA}^2$  away from any of its closest neighbors. The volume of each subunit was defined by the van der Waals radii of the constituent atoms along with a  $1.4\text{\AA}$  addition accounting for water. The density of the dots is high 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  $\sim 1.7\text{\AA}$ .



**FIGURE 3.2: Calculating % holes and overlaps.** An atom resolution model (**A**; here, a levivirus–PDBID 1qbe—was used for its high overlaps and holes) taken from the VIPERdb [8] is projected onto a sphere of mean radius (**B**). A “dot matrix” is then cast (shown independently in **C**) that is then used to find the extent of subunit-subunit overlaps (**D**; calculated from “double occupancy” dots that are present in two subunits) and holes (**E**; calculated from “zero occupancy” dots that are present in no subunits). Those capsids displaying low % holes and overlaps are considered as potentially tilable by a single tile (i.e. monohedrally tilable).

The last test is to see whether the “tiles” or subunits are structurally invariant.

### 3.1.2.3 Test for capsid “monohedrality”

Assuming that the capsid is tilable, monohedral tilability is found when the *shape* of each tile (or subunit) is the same (or congruent)<sup>11</sup>. 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

<sup>11</sup>Note that the subunits don’t have to be related by symmetry to possess the same shape.

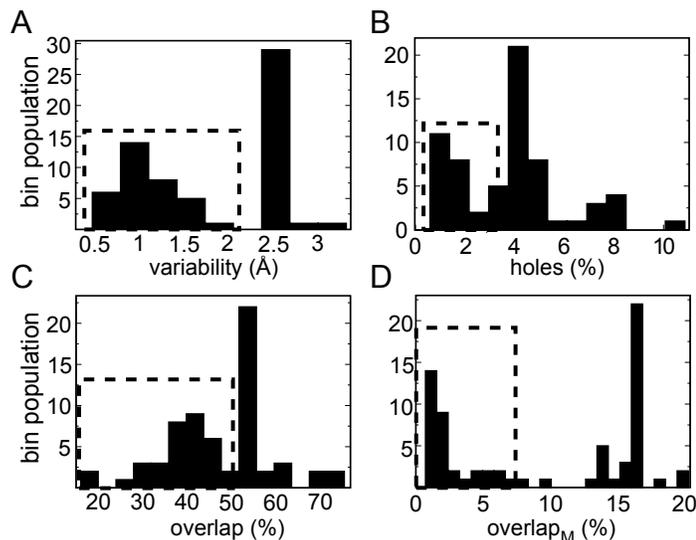


FIGURE 3.3: There are a large number of spherical capsids, highlighted by the dashed rectangles, that possess *at least one* of the three requirements of monohedral tilability.

not localized conformational changes which alter the inter-subunit interactions in an otherwise structurally rigid subunit, e.g., the order-disorder transitions in the tomato bushy stunt virus capsid (reviewed in [14]). Those structural changes may be manifested in the tiling as subunit-subunit dihedral angle changes.

### 3.1.3 Tilability of natural spherical capsids

It is immediately evident from the histograms (Figure 3.3)<sup>12</sup> that there is one group of capsids (highlighted by the dashed rectangles), where *at least one* of the three prop-

<sup>12</sup>**Note on method:** In Figure 3.3, the four histograms probe the following four properties: (a) monohedrality (subunit variability within a capsid measured by an averaged RMSD value in Å), (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 (which was calculated by first shrinking each pruned subunit (“Pruned”, here, means that the those amino acids within a subunit that undergo order-disorder transitions [14], which partake in the modifying of one interface, are ignored in the assay) by a scaling factor of 0.83 and then calculating the percentage overlaps without the 1.4 Å addition to each atom radius). 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 most well behaved 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.

erties of monohedral tilability are possessed (in all of the metrics used, low values indicate that the capsid possess little structural variability, negligible holes or negligible overlaps). The next question is: do some capsids possess *all three* properties (making them representable by monohedral tilings)?

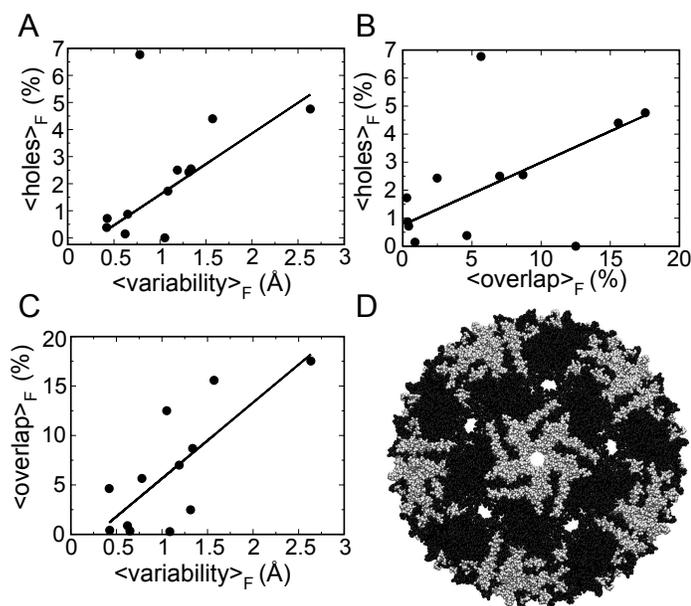


FIGURE 3.4: (A,B,C) The dependence of one of three requirements for monohedral tilability is plotted versus each of the others using family averaged values. The solid lines are added to emphasize the trends. (D) An example of a capsid belonging to the Leviviridae family (PDB ID: 1mst), which violates all of the requirements of monohedral tilability, as indicated by excessive overlap of black and white subunits and large holes at five- and pseudo six-fold symmetry axes (Subunit coloring: A,B: black; C: white).

This question is answered in Figure 3.4. The average % overlap<sub>M</sub>, % 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 Figure 3.4A, B and C, respectively. 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, the weak exceptions being the families tetraviridae and hepadnaviridae.

It is evident that there are capsids (belonging to families in group 2 in Tables 3.1) that present either high structural variability, subunit-subunit overlaps (tetraviridae), holes (hepadnaviridae), or all three characteristics (leviviridae and polyomaviridae) and prevent them from being represented as monohedral 2D bound tilings, or canonical capsids. A levivirus has been represented in Figure 3.4D 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 Tables 3.1) that possess all three qualities of monohedral tilability and “reside” within all the dashed-line boxes in Figure 3.3. These capsids, may be represented by bound monohedral tilings that we call canonical capsids.

Although the bimodality of the histogram distributions in Figure 3.3 (into group 1 and group 2) is evident, there *is* some group 1/group2 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.

The next section hopes to show that predictions made on the *platonic* group 1 virus capsid–discrete mathematical models or canonical capsids–can, indeed, be related back to the capsids belonging to families in group 1 of Tables 3.1.

►► The second half of this chapter will ask the following question: if the capsid was made up of “Lego blocks”, what would be the optimal properties of each block?

We will show that our predicted subunit design (Section 3.2.4) is remarkably similar to those that we see in a diverse number of natural viruses (Section 3.4), which has interesting implications regarding design and evolution.

## 3.2 Characterizing the subunit shape

Here, we will attempt to characterize the types of shapes available to the canonical capsid *subunit* by sequentially answering two more tractable questions (simply because it felt like the easiest way to go about doing so). First, we will find the number of *edges* (denoted in symbolic form as  $\sigma$ ) that the canonical capsid prototile may possess<sup>13</sup>. Then we search for specific *tilings* (and their allowable shapes) that are in accordance with the canonical capsid definition and the allowable  $\sigma$ s (read on, both activities are surprisingly easy).

### 3.2.1 I. The number of sides ( $\sigma$ )

To answer the first part of the question (number of edges allowed), we need a subunit model that is “malleable”, i.e., one that can readily change its edge number and shape (remember we are asking about edge number,  $\sigma$ , not shape). It easily follows that topology and graph theory—that deals with surfaces and sets of connections, and not

---

<sup>13</sup>“Prototiles” are *tiles that are used as building blocks for tilings*. If we have a monohedral tiling (Section 3.1.1.3), then we will have only one prototile repeated many times that need to be used to create the tiling. Still, the number of prototiles that may independently be used to assemble a canonical capsid are not known, and we are out to seek that information.

geometry—will be used to answer the  $\sigma$  question. Before all that, we need to formulate a representation of the subunit that enables these fields to be utilized.

### 3.2.1.1 Representing the subunit

Figure 3.5A shows a representation of the prototile with one quasi-equivalent vertex (circled), two straight edges, and one curved edge. The curved edge may possess any

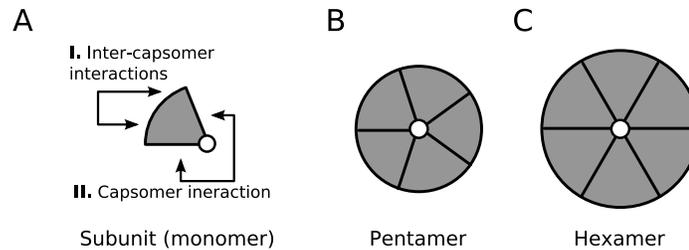


FIGURE 3.5: An abstract representation of a canonical capsid subunit is represented here in subunit form along with the types of subunit-subunit interactions (A) and in capsomer (hexamer and pentamer) arrangements that are formed from type II interactions (B,C). The final capsid is composed of 12 pentamers and  $10(T - 1)$  hexamers that interact with each other via type I interactions [12]. 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).

number of 3-valent vertices (allowing this structure to have an arbitrary number of edges/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 (Figure 3.5B,C), each quasi-equivalent vertex may be either five or six valent. The number of vertices and edges ( $\sigma$ ) may be dependent on various properties of the capsid such as the size or  $T$  number.

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

### 3.2.1.2 Bringing out the math

Monohedral tilings (described by edges, vertices and faces) describing canonical capsids must abide by Euler's polyhedral formula<sup>14</sup>, which shows that various elements within polyhedra may be related to each other in a predictable manner. Specifically, the number of vertices ( $V$ ), edges ( $E$ ), and faces ( $F$ ) of a such a polyhedron may be related through the following equation:

$$V - E + F = 2 \quad (3.1)$$

Now, we will try to obtain a relationship between the allowable number of interactions/edges per tile ( $\sigma$ ) and the variables in Equation 3.1.

Since we know that the number of faces corresponds to the number of subunits in the canonical capsid (equaling  $60T$  subunits per capsid with triangulation number  $T$ ; see Section 2.1.5), we have

$$F = 60T \quad (3.2)$$

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

$$E = F\sigma/2 = 30T\sigma \quad (3.3)$$

---

<sup>14</sup>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 [36]). 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.

The Caspar and Klug rules of capsid assemblies (Section 2.1.5) shows that there are twelve 5-valent vertices and  $10(T - 1)$  6-valent vertices [12] within the canonical capsid (we call these the centroid vertices, which are  $V_C$  in number). So, the number of centroid vertices

$$V_C = 10(T - 1) + 12 = 10T + 2 \quad (3.4)$$

We know that each prototile must have exactly one centroid vertex<sup>15</sup>. 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 in the *polyhedron* is equal to the number of subunits ( $60T$ ) times the number of trivalent vertices per subunit  $(\sigma - 1)$  divided by the number of subunits that share each trivalent vertex (3), i.e.,

$$V_R = 60T(\sigma - 1)/3 = 20T(\sigma - 1) \quad (3.5)$$

Finally, from Equations 3.4 and 3.5, we have the total number of vertices

$$V = V_R + V_C = 2 + 20T\sigma - 10T \quad (3.6)$$

Substituting Equations 3.6, 3.3, and 3.2 in Equation 3.1, we get

$$(2 + 20T\sigma - 10T) - (30T\sigma) + 60T = 2 \quad (3.7)$$

Further reducing this equation, the  $T$ s cancel out and we get

---

<sup>15</sup>This comes from the knowledge that pentamers and hexamers are *coordinated* at the center.

$$\sigma = 5 \tag{3.8}$$

### 3.2.2 A capsid invariant

All the equations above converge to this fun little form in Equation 3.8. And when I first had the pleasure of canceling the  $T$ s out of Equation 3.7, and it was exciting. Why? Because a messy equation turning into a simple one is a beautiful thing. Important for capsid design, though, is that we obtained an invariant geometric characteristic ( $\sigma$ ) that is independent of the  $T$  number. This indicates that a subunit may be modified to assemble into capsids of non-native  $T$  numbers. As a corollary, it also means that capsid subunits of one shape can potentially assemble into capsids of any size.

►► If capsid size is not modulated by general subunit *shape*, then how is this size modulated? That exciting question is tackled in the next chapter.

So far, we established that the subunit must be five sided<sup>16</sup>. The next part of the puzzle is in asking what shapes (or tilings) can this five-sided subunit have?

### 3.2.3 II. The shapes available to a five-sided subunit

As per the definition of the canonical capsid, any prototile that can tile a bound canonical capsid should, by the rules of quasi-equivalence, also tile a two dimensional hexavalent lattice (see Figure 2.2 and Refs. [12] and [37]). So, our second query is simplified to: how many five-sided prototiles may assemble into a hexavalent lattice?

Fortunately, the tiling of 2D surfaces by convex pentagons has been a subject studied at least since the early twentieth century. In the early 1980s, Grünbaum and

---

<sup>16</sup>We also rationalize this notion with the help of symmetry in Section 3.5.

Shepherd enumerated a list of thirteen known convex pentagonal edge-edge tilings, and since then, one more has been added to that list [38, 39]. We searched the resulting catalogues as capitulated by Sugimoto and Ogawa [38] for convex pentagonal prototiles that can assemble into a hexavalent lattice, and found only one tiling to fit that criteria (known in [38] as *Type 5*; shown here in Figure 3.6A). 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.

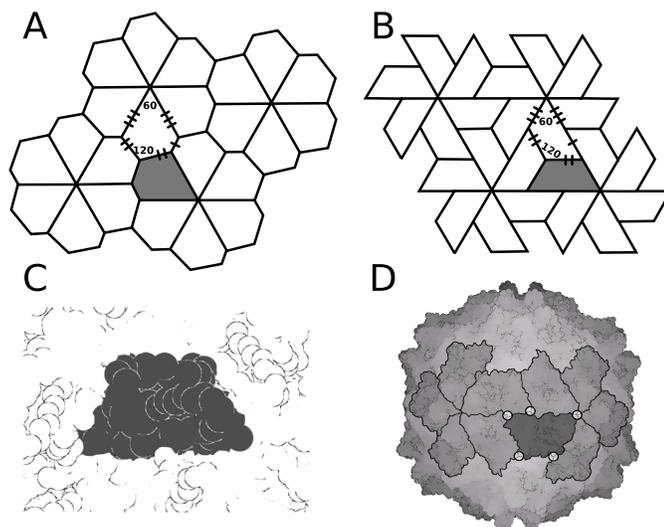


FIGURE 3.6: *Pentagonal type 5 tilings*. *Type 5* tiles (e.g., **A** and **B**) are the only known five sided tiles that possess six-valent vertices. The interacting edges of the prototile in both **(A)** and **(B)** have been marked with one, two and three dashes. Such edges interact with identically 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).

That topology constrains subunit form is not a new idea, and Thompson’s “Growth and Form” stated a similar idea with respect to using shapes to create a polyhedron<sup>17</sup>.

In their famous 1956 paper, Crick and Watson cited this statement as “essentially a

<sup>17</sup>Thompson stated that “the broad, general principle that we cannot group as we please any number and sort of polygons into a polyhedron but that the number and kind of facets in the latter are strictly limited to a narrow range of possibilities”.

topological one” [9]. So, then, the surprise in our work is not that the shape of the capsid subunit appears to be constrained, but that the topological constraint appears to be strong, and only a small number of shapes may be allowed in the canonical capsid world.

Most importantly, the tiling described in Figure 3.6A is equivalent (combinatorially) to the trapezoidal tiling seen in Figure 3.6B, which represents a 2D projection of the trapezoid structure seen in nature (e.g., Figures 3.6C and D). Lets take a detour to emphasize the importance of this sentence.

►► Mathematically, the trapezoid is the best solution to the canonical capsid subunit shape.

### 3.2.4 The ubiquitous trapezoid shape emerges

In the late 1960s, multiple labs were attempting to elucidate the first capsid structure ever. The race for this distinction was fervent, and, at the end, the first high-resolution capsid structure (that of the tomato bushy stunt virus) was elucidated by Steve Harrison’s lab [40]. Michael Rossmann’s lab came in at a close second with the southern bean mosaic virus [41]<sup>18</sup>. Often, coming second is not that nice, but this time, the consolation prize was as good as winning, since Michael Rossmann made an interesting discovery (although there were hunches before). He made the observation that the protein subunit from both viruses shared a remarkably similar structure<sup>19</sup>. Since then, that structural motif has been seen repeatedly in plant, animal and bacterial viruses,

---

<sup>18</sup>Incidentally, two members of my thesis committee—Art Olson and Jack Johnson—were present as post doctoral scientists in each of those labs, and are co-authors on those historic papers!

<sup>19</sup>That eight stranded beta barrel motif, otherwise known as the jelly roll, has been associated with Rossmann ever since.

which makes it one prolific shape. It is also the same general shape that our theory predicts for canonical capsids. But why do we see this shape more than any other?

### 3.3 Divergence or convergence?

*Divergent evolution* is an event where a variety of “species” evolve from a common ancestor (i.e., they *diverge* from the common ancestor). During this event, various features of the ancestor (such as a backbone for vertebrates) are retained in the current-day species. In the species world, if one sees two species that have similar traits, they probably had a common ancestor (this is for various reasons, but large system size is the one in focus in Section 1.3.1). In the species world, it is unforeseeable that two highly similar organisms could possess similar traits unless they diverged from a common ancestor, and hence, divergent evolution is the theory that most satisfactorily explains the existence of highly similar forms.

So, going back to these viruses, looking at the trapezoidal subunit’s ubiquitous presence<sup>20</sup>, one could be tempted to state that all of them descended from a common ancestor. In fact, divergent evolution *is* the running explanation to the trapezoidal structure, i.e., it is believed that all these capsids displaying trapezoidal subunit shapes arose from a single proto-viral strain [42]. This notion has been especially plausible because, so far, it would have seemed inconceivable that distinct evolutionary lineages could ever converge on a single protein shape (that mode would be described as *convergent evolution*).

---

<sup>20</sup>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).

However, here, we show that the bisected-trapezoid is the only subunit shape (among other type 5 tilings) that is allowable within canonical capsids<sup>21</sup>. 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). This idea is made more plausible if we recognize that viruses display (1) high genome mutation rate (i.e. high “sampling”), (2) low system complexity (compared to other biological phenomena) and (3) high population doubling rate (see Section 1.3.2 for more discussion).

### 3.4 Canonical capsids *do* represent group 1 capsids

The last question that we ask is: which virus capsids possess these trapezoidal shapes? One possibility is that these subunit designs are sprinkled randomly within the capsid world—among both “tilable” group 1 capsids and others (Table 3.1). That would be bad, since our predictions should be stringent only for the group 1 canonical capsids ( that display few overlaps, holes and little subunit-subunit variability).

Fortunately, we discovered that *all* capsids present within group 1 families described in Tables 3.1—the capsids representable by monohedral tilings (or canonical capsids)—also possess trapezoidal shapes. This indicates a strong relationship between our predictions on our mathematical models and the features seen in natural capsids,

---

<sup>21</sup>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/monohedral (all of these features are acceptable from a free energetic point of view).

which indicates that the canonical capsid may, in fact, be used to represent and study capsids belonging to group 1 families of Tables 3.1. 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 [43].

### 3.5 Rationalizing the need for a five-sided subunit

One could look at the need for a five-sided subunit intuitively from a symmetry-oriented standpoint (please refer to Section 2.2.4 for a review on mirror symmetry and chirality). There are two 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 2, 3 and 5-fold axes of rotation (see Section 2.2.2), they differ in the number of mirror planes they possess; the full icosahedral symmetry group possesses fifteen mirror planes while the rotational icosahedral symmetry group possesses none. As chiral structures can not possess mirror symmetries (see Section 2.2.4), due to the chirality of biological macromolecules, it follows that only the chiral icosahedral symmetry may exist.

Figure 3.7 represents versions of the three, four and five sided prototile in its hexameric form (a), which is a precursor to the canonical capsid, and the prototile itself (b). It is immediately evident that the number of possible mirror symmetries (shown as dashed lines) diminishes to zero at  $\sigma = 5$ . So, due to the molecules chirality,  $\sigma = 5$

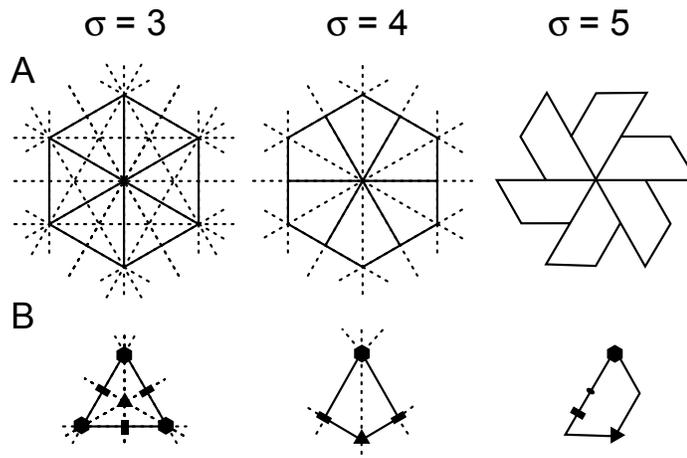


FIGURE 3.7: **Mirror symmetries preclude 3 and 4 sided subunits** (in canonical capsids). The relationship between mirror symmetry elements and the number of edges ( $\sigma$ ) per hexavalent tile. This image illustrates trigonal, tetragonal and pentagonal tiles in (A) the hexavalent form and (B) as a single tile (along with the 2, 3 and 6-fold 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.

is the only viable option for virus capsid subunits; and from  $\sigma = 5$ , the trapezoid of the tiling is automatically obtained.

### 3.6 Breaking the canonical rules

Capsids that possess subunit-subunit holes, overlaps or structural variability (i.e., group 2 capsids) were found to possess, largely, non-trapezoidal shaped subunits. Most notable among these are the polyomaviridae family viruses [44, 45] that not only break the rules of monohedral tilability, but also break the otherwise longstanding “rules” of quasi-equivalence (although at the capsomer level, these rules are still maintained loosely). We propose that the breaking of Caspar and Klug’s rules of quasi-equivalence

[12] are only possible with the use of (1) chemically distinct subunits or (2) subunits that grossly break the rules of monohedral tilability.

### **3.7 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 quasi-equivalence by displaying only pentamers and no hexamers [27–30]. Specifically, they showed that capsids belonging to the polyomaviridae family must be represented as bound tilings formed from *two* distinct subunit shapes or tiles [27, 30], i.e., capsids from this family can not be represented by monohedral tilings, which corroborates the classification of these capsids as “group 2” in Tables 3.1. Importantly, Twarock and colleagues showed that one may characterize the “assembly pathways” of these tilings using mathematical methods [28, 29]. 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 quasi-equivalence–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.

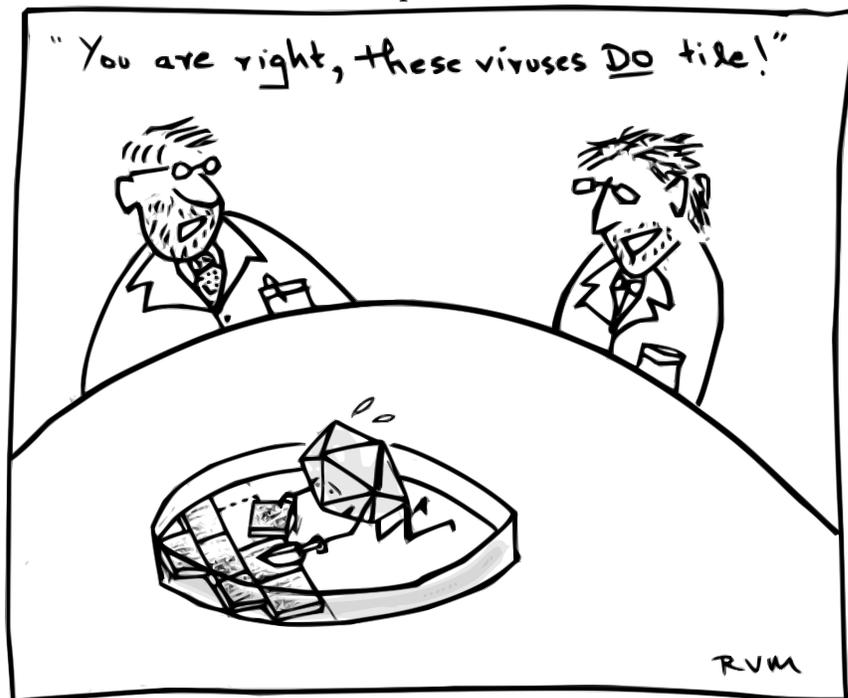
## 3.8 Conclusion

★ Canonical capsids, described in Section 3.1.1.3 as monohedral tilings, represent a large number of capsids (Section 3.1.3).

★ Simple math predicts that, for canonical capsids, the trapezoidal subunit (and its variants; see Section 3.2.4 and Figure 3.6) is the only available subunit shape. This prediction was shown to be true for all canonical capsids in nature (Section 3.4), which indicates that independently evolving capsids may plausibly converge on this universal subunit design (Section 3.3).

★ Most importantly (at least for the next chapters), the correspondence between our models and real capsids lends credence to usefulness of the bound monohedral tiling model of capsids (the “canonical capsid”).

(pew)



... Next chapter ...

Geometry rules

Hexamers dictate capsid size

Viruses swell

(and we get more confidence in our models)

## Chapter 4

# Geometric properties of the virus capsid

Leaving the capsid subunit, our focus will now be turned to analyzing the structure and mechanics of *clusters* of subunits. Virus capsids come in various sizes (Figures 2.1 and 2.2). Still, single virus strains often need to form capsids of *specific* sizes in order to complete their virus life cycles. Despite its importance in pathogenicity, little is known regarding the determinants of capsid size. Still less is known about exactly which capsids can undergo maturation events such as buckling transitions—post-capsid-assembly events that are crucial to some virus strains.

In this chapter, we will show that the exclusive determinant of capsid size is *hexamer shape*<sup>1</sup>. This conclusion arises from considering the dihedral angle patterns within hexamers belonging to natural canonical capsids and geometric capsid models (deltahedra). From simple geometric models and an understanding of *endo angle*

---

<sup>1</sup>The idea is that specific hexamer shapes can only be assembled into specific capsid sizes.

*propagation* discussed here, we then suggest that buckling transitions may be available only to capsids of certain size (specifically,  $T < 7$  capsids are precluded from such transformations) and that  $T > 7$  capsids must require the help of auxiliary mechanisms for proper capsid formation; These predictions, arising from simple geometry and modeling, are backed by a body of empirical evidence, further reinforcing the extent to which the evolution of the atomistically complex virus capsid may be principled around simple geometric design/requirements.

Specifically, this chapter asks two questions about the virus lifecycle:

- ✓ How do subunits form into capsids of specific sizes? (the size specificity problem) and
- ✓ Which capsid can undergo these special structural transitions called buckling transitions?

This chapter is derived from the following paper: Mannige, R.V. and Brooks III, C.L. (2009) Geometric considerations in virus capsid size specificity, auxiliary requirements, and buckling. *Proceedings to the National Academy Science U.S.A.*, 106(21):8531-6.

## 4.1 Introduction

A large number of human and crop-infecting viruses are protected by spherical capsids (shells) of various sizes that are primarily made up of self-organizing protein subunits [46, 47]. Caspar and Klug's seminal paper on quasi-equivalence [12] explained how an infinite range of capsid sizes can be "constructed" by combining  $60T$  subunits or twelve pentamers (five-valent subunit clusters) and a variable number of hexamers

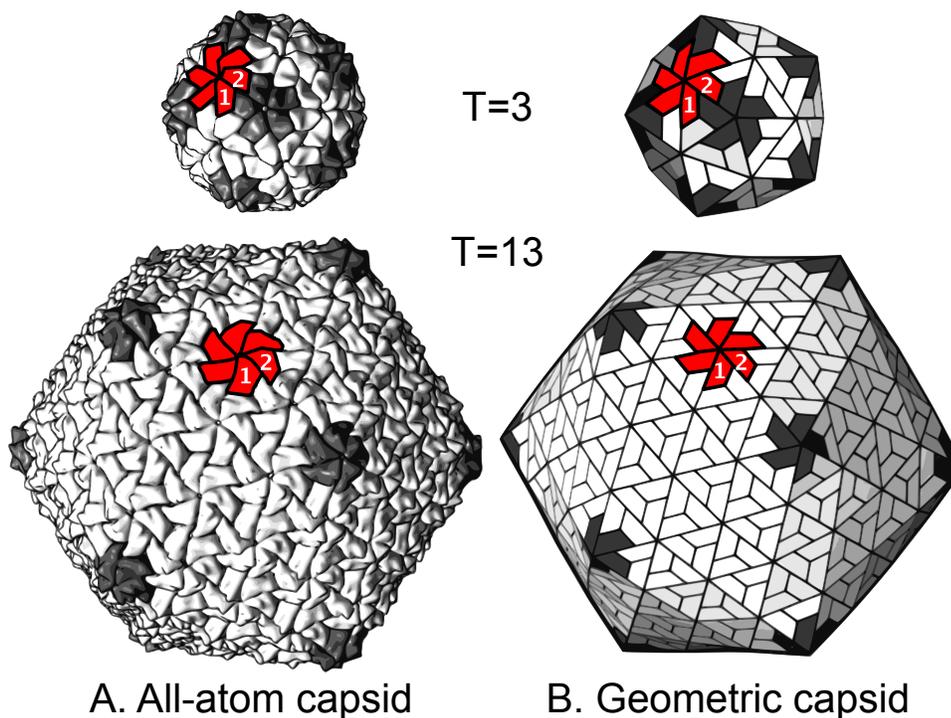
$(10 \times (T - 1))$  into a closed spherical shell ( $T = 1, 3, 4, 7, \dots$  and is the *triangulation number* [12] described in Section 2.1.3).

From the range of possible sizes, generally, subunits from specific viral strains assemble into capsids of specific sizes; the inability to form those native sizes is believed to result in the loss of infectivity. For example, the sobemovirus and birnavirus capsids [48, 49] shown in Figure 4.1A are known to be pathogenic primarily in their native  $T = 3$  and  $T = 13$  capsid forms (or sizes), respectively. Despite its importance in pathogenicity, our picture of capsid size-specificity is incomplete. Here, we are interested in the structural features (constraints), if any exist, that differentiate between capsids of different sizes (capsid design criteria). An appreciation of these concepts is pressing from a nanotechnological perspective (for the rational design of artificial scalable assemblies that build on current practices, such as in the use of protein fusion and symmetry properties by Padilla et al. [50]) and a therapeutic perspective (to impede the formation of infective native capsids).

The size-specificity puzzle gets more interesting given the theoretical evidence (derived in Chapter 3) that a single subunit shape (the trapezoidal prototile) possesses the ability to tile all of the allowed canonical capsid sizes ( $T = 1, 3, 4, 7, \dots$ ) [1], which is backed by evidence of a ubiquitous trapezoidal subunit shape seen in nature (discussed in [42]). In these situations, the differences between capsids of different sizes will be seen within the capsid's subunit-subunit dihedral angles<sup>2</sup>, i.e., size-specificity within canonical capsids (defined in Section 3.1.1.3) may be manifested not in the subunit

---

<sup>2</sup>For example, the average dihedral angle value per capsid will tend towards  $180^\circ$  as we proceed to larger and larger capsid sizes.



**FIGURE 4.1: Two capsid sizes.** In (A), two natural canonical capsids (the  $T = 3$  sobemovirus and  $T = 13$  birnavirus capsids with PDBIDs 1smv and 1wce respectively) are shown to emphasize that spherical capsids come in many sizes that are composed of 12 pentamers (dark grey) and  $10 \times (T - 1)$  hexamers [12]. We use geometric models (B) as platonic capsid representations for the characterization of structure and function. In each capsid, a single hexamer (colored red) along with two subunits (“1” and “2”) are marked to emphasize the structural correspondence between all-atom and geometric capsids.

shape itself but in the angles at which the generally rigid subunits interact within the capsid.

In the following sections, we attempt to show that the exclusive determinant of canonical capsid size is hexamer *shape* as defined by the internal subunit-subunit dihedral angles within the hexamers. We then use knowledge of “endo angle constraints” (defined here) to predict that only capsids of specific sizes ( $T \geq 7$ ) possess the potential to undergo true buckling transitions. Interesting inferences on the requirement

of auxiliary proteins in large capsids are also discussed in the context of hexameric flexibility.

Next, we look at subunit-subunit dihedral angles in hexamers.

## 4.2 Hexamer shapes encode for capsid size

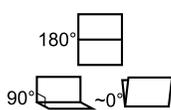
To understand capsid size specificity, we chose to focus on hexamers and the effect of neighboring pentamers on their shapes for various reasons: From a geometric perspective, capsids of different sizes (but formed from similarly shaped subunits) must possess identical pentamers<sup>3</sup>, which has also been shown to be true for both cryo-EM [51] and x-ray structures<sup>4</sup>. Secondly, hexamer structure in some capsids is known to be influenced by the presence of neighboring pentamers [51, 52], indicating that the arrangement of neighboring pentamers may be important for defining hexamer shape. Also, if the subunits in the hexamer are expected to be the same, then the hexamer's six dihedral angles<sup>5</sup> must define hexamer shape.

Comparing dihedral angles between subunits involves defining subunit planes (and then comparing the angle between adjacent planes), which is an imprecise endeavor as the subunit is a three dimensional molecule with a rough and complicated atomic

<sup>3</sup>Discussed in Section A.1.

<sup>4</sup>Crystallographically, this is evident when comparing pentamers appearing in capsids of two sizes ( $T > 1$  and  $T = 1$ ) that are formed from chemically identical subunits, e.g., in the birnavirus (PDBIDs:  $T = 13$ : 1wce;  $T = 1$ : 1wcd), alfalfa mosaic virus (PDBIDs:  $T = 3$ : 1js9;  $T = 1$ : 1yc6), and sesbania mosaic virus (PDBIDs:  $T = 3$ : 1smv,  $T = 1$ : 1x36).

<sup>5</sup> **What is a dihedral angle?** It is the angle between any two planes. Examples are given below.



If you have a sheet of paper and draw a line on it, the dihedral angle between the two halves is  $180^\circ$  or planar. Fold the paper across that line such that the two halves are at right angles to each other and we have a dihedral angle of  $90^\circ$  at that line. Fold the paper onto itself, and the angle is  $0^\circ$ .

surface<sup>6</sup>. Instead, we looked at how *similar* the pentameric (defined “endo”) dihedral angle is to each of the hexameric angles within a capsid. For any hexamer within capsids possessing highly uniform subunit structures (i.e., strictly canonical capsids [1]), this can be done simply by structurally aligning a *pair* of adjacent pentameric subunits to each of the six pairs of adjacent hexameric subunits (more in Materials and Methods). Each pair-pair structural alignment results in one root mean square deviation (RMSD) value, which is low if the angles associated with the pairs are similar (and 0 if the pairs possess identical angles).

In each capsid studied, for every unique hexamer in a distinct environment ( $T = 13$  capsids have two unique hexamer environments, while  $T = 3, 4$  and  $7$  possess just one), we obtained six RMSD values (numbered 1 through 6 in counter clockwise fashion starting with an angle closest to the pentamer) represented as lines (one for each unique hexamer) and grouped by  $T$ -number in Figure 4.2A (shown separately for each capsid in Figure A.2). Excepting the  $T = 13$  capsid, which possesses two unique hexamers (labeled as “hexamer 1” and “hexamer 2”), each line in Figure 4.2A is obtained from distinct natural canonical capsid structures (described in Materials and Methods). The qualitative groupings of the lines in this figure suggest that hexamers exist in variously puckered hexamer shapes that are size or  $T$ -specific. For example, all hexamers from  $T = 3$  and  $T = 4$  canonical capsids appear to display characteristic “ruffled” and “wing” shapes, respectively, displayed geometrically in Figure 4.3C (that correspond to previously described trimer of dimers and dimer of trimers [53, 54], re-

---

<sup>6</sup>Also, many capsid subunits “display” protruding domains on the capsid’s surface (e.g., the **P** domain of the Tomato Bushy Stunt Virus, PDBID: 2tbv), making the choice for a suitable generalized plane even harder.

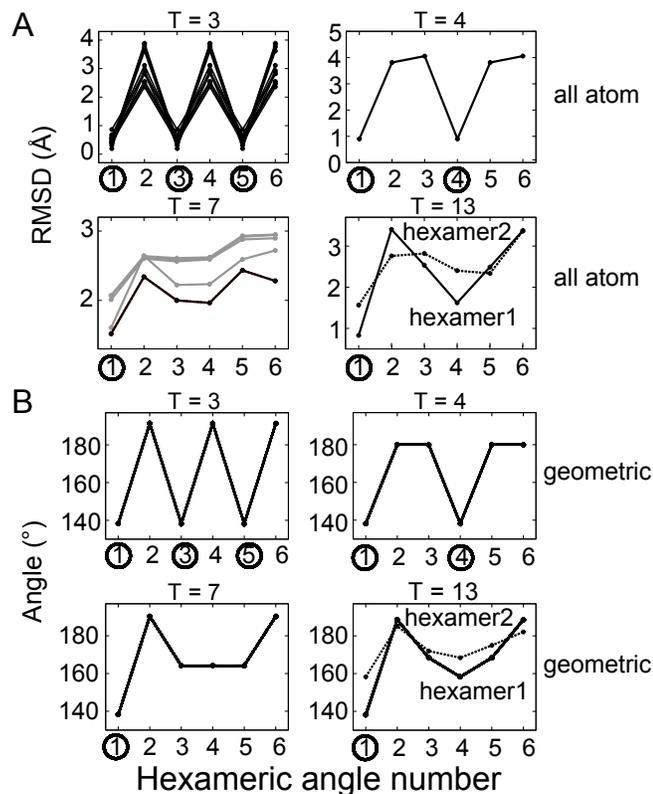


FIGURE 4.2: **Hexamer shape is specific to capsid size.** (A) shows the extent to which hexameric dihedral angles (numbered 1 through 6 on the x-axis in all graphs) found within natural capsids resemble the endo angle found within the pentamer (angle similarity is proportional to the RMSD). It is evident that hexameric angles adjacent to pentamers (with numbers circled on the x-axis) are consistently endo-like, which gives rise to the concept of *endo angle propagation* (read text). Furthermore, *ab initio* (geometric) models were used to obtain accurate hexameric dihedral angle values (B), which reflect the patterns seen in (A). Both sources (A,B) indicate that the shapes available to the hexamer is strongly constrained by the size of the capsid. An expanded version of (A) is available in Figure A.2

spectively). Also, all hexameric angles (circled in the x-axis in Figure 4.2A) adjacent to pentameric “endo” angles are also *endo like* in nature, as indicated by the low RMSD values, which is an important outcome of the pentameric *endo angle constraint* on hexamer shape discussed further on.

✓ Hexamer shape is capsid size specific and is primarily constrained by the neighboring arrangements of pentamers.

Figure 4.2A, however useful, cannot be used in making quantitative observations on hexameric geometries that would be required from a capsid design/nanotechnology perspective (since an RMSD does not provide us with angle values, it is only an angle similarity metric). For that, computational models of canonical capsids (deltahedra, described in Materials and Methods) were built for  $T$  numbers 1, 3, 4, 7 and 13<sup>7</sup>. The dihedral angles present within model hexamers are plotted in Figure 4.2B, also arranged as per capsid size. It is clear that these dihedral angle patterns closely resemble those seen in nature for each available capsid size (Figure 4.2A). This on its own is interesting because these models were obtained from independent *ab initio* methods (obtained from non-dimensional deltahedron graphs embedded in three dimensional space with no application of icosahedral symmetry) but still display natural canonical capsid properties (hexamer shape), reinforcing the natural capsids' geometric/mathematical nature. More relevant, however, is that geometric constructions *independently* reiterate hexamer size-specificity. Additionally, even in the models, it is evident (partly because of the geometric construction itself) that those hexameric angles shared with pentamers are pentameric (or endo like) at  $\sim 138.19^\circ$ , independent of size (or  $T$ ). Section A.1 shows that this is a mathematical result of quasiequivalence and the monohedral tilability of the capsid.

✓ Hexamer shapes described in Figure 4.2B may be used in creating size-specific artificial assemblies.

<sup>7</sup>Note that deltahedra have been previously discussed with respect to spherical capsids, particularly *Fig. 8* in [12] and *Fig. 3* in [52]; however, in both studies, the deltahedra were conceptual tools, and could not be readily related to natural capsid arrangements; only from recent observations of monohedral tilability [1] discussed in Chapter 3 can we now represent a large number of natural capsids by deltahedra in a structurally meaningful manner.

### 4.2.1 Endo angle constraints

From our analysis of both the all-atom capsid structures and geometric models, we can surmise that hexameric dihedral angles are affected by the presence of adjacent pentamers. Figure 4.3A represents a canonical capsid subunit (described in [1]) with its interaction types that give rise to all possible capsid sizes, and B represents a pentamer-hexamer cluster present in  $T > 1$  capsids. It is evident, if all subunits within a capsid retain similar shape and size, that the angle within the pentamer will be propagated into the adjacent hexamer (indicated by the arrow). This we call the *endo angle constraint*. From this it becomes evident that hexameric shapes (Figure 4.3C) must be specific to capsid size. This is a natural progression of the endo angle constraint on account of shifting positions (and numbers) of neighboring pentamers around the hexamer (an effect that is corroborated by discussions on empirical [51] and theoretical [52] bacteriophage models<sup>8</sup>).

We expect that the dihedral angles within the remainder of the hexamer (the “unconstrained” angles we call *exo* or “*x*”) must also be indirectly constrained by endo angle propagation, as they must accommodate values suitable to the distribution of the preset endo angles, i.e., the number of endo angles present within a hexamer will be important in determining the possible shapes available to the hexamer.

---

<sup>8</sup> [51] dealt with pentamer polymorphism within a single capsid, and Moody [52] reasoned that the hexamer shape was modified by the distortion of pentamers due the projection of the pentamer onto the icosahedral insphere (see “hexamer rectification” in [52]). Although very creative and useful qualitative rationalizations of some cryo-EM structures, these rationalizations are clearly not applicable to canonical capsids with uniform subunit shapes.

- ✓ Pentamers appear to impose endo angle values onto adjacent hexameric dihedral angles.
- ✓ This imposition dictates that hexamer shape must be capsid size (or  $h,k$ ) specific.

### 4.2.2 Considering larger capsid sizes

As we approach larger capsid sizes ( $T > 7$ ), the number of hexamers in unique environments will increase. We propose that capsids of all sizes may be created from a small repertoire of distinct hexamer shapes. Early work showed that capsids may be separated into three classes distinguishable by distinct size-specific capsid morphologies (obtained from symmetry considerations in [11], paper models in [12], endo angles in Section 5.3.2 and Ref. [1]), e.g., capsids with  $k = 0$  (where  $T = 1, 4, 9, 16, 25\dots$ ) belong to *class 1* in Figure 5.3 (and *class A* in *Fig. 10* in Ref. [11]) are most icosahedral in morphology.

As a harbinger of the chapter to come, we will expand on this class system, and hypothesize that capsids belonging to the same class will possess conserved/common hexamer shapes, and so, the rational modification of a capsid's size within a class will be easier than inter-class size conversions. This explains why  $T = 4$  capsid subunits, that form “wing” shaped hexamers (Figure 4.3C), once mutated, are able to assemble exclusively into other sizes within the same morphological class ( $T = 1, 4, 9, 16, 25,$  and  $36$ ) [55]. These interchangeability rules explain how capsids of various sizes may have been sampled from a relatively simple set of capsomer building blocks, leading to a range of capsids seen today.

### 4.2.3 $T > 1$ to $T = 1$ capsid transformations

We find that capsid models of any size possess pentameric angles equaling  $\sim 138.19^\circ$ , the same as internal angles within an icosahedron and that seen within  $T = 1$  pentamers. Considering that no explicit icosahedral symmetry was enforced onto the building of the models, this is not an expected result, as a pentamer (collection of five valent plates) possesses a configurational degree of freedom (and could therefore possess non-similar angles). This indicates that, even at the most basic geometric level, subunits fated to form  $T > 1$  capsids may possess enough information to assemble into  $T = 1$  capsids, especially if subunit-subunit angles specific to hexamers are prevented from forming, an effect visible in both canonical [49, 56] and non-canonical natural capsids [57, 58]. Consequently, *all* subunits evolved to form  $T > 1$  capsids may possess the potential for  $T > 1 \rightarrow T = 1$  transformations in specific conditions.

✓ All capsid subunits may possess the ability to form  $T = 1$  capsids.

### 4.2.4 Implications for anti-capsid therapies

Our results suggest that rational control of hexamer shape may allow for the redirection of native capsid subunits into non-native/non-infectious forms, allowing for the development of non-native but industrially useful assemblies and, more importantly, allowing for rational/combinatorially directed anti-viral drug design. An example of such hexamer shape modification is by the binding of organic molecules to specific intra-hexamer subunit-subunit interfaces (e.g., the molecule HAP1 that modifies  $T = 4$  capsid assembly in the Hepatitis B virus [59]). The current use of organic molecules in

controlling capsid disassembly (discussed in [42]) and assembly (e.g., [59–62]) provides a possible platform to commence the rational search for such molecules that modulate capsid size via the modification of hexamer shape.

### 4.3 Endo angles and buckling transitions

Some capsids undergo buckling transitions, where the capsid, once assembled, undergoes a change in morphology from being more spherical to a more faceted (or more “icosahedron-like”) form [63, 64]. For capsids that undergo such transitions (distinguished from capsid “swelling events” below), this change in morphology is crucial to the continuation of the virus life cycle. In what follows, we attempt to validate a hypothesis that emerges from our understanding on endo angle constraints. Let us consider hexamers extracted from deltahedra for a range of sizes or  $T$  numbers (Figure 4.4A). Here, the solid lines represent rigid edges, equilateral triangles represent flat subunits, and  $\mathbf{P}$  marks the dihedral angle that is shared with a pentamer and hence endo constrained at  $\sim 318.19^\circ$  (the constraint is depicted as dashed lines that prevent specific angles from changing). We hypothesize that buckling transitions should be possible only in  $T \geq 7$  capsids, where the total number of endo angle constraints per hexamer (given by  $6/[T - 1]$ , which is easily derivable<sup>9</sup>) is one or lower. From a brief analysis of the graphs in Figure 4.4A, it is evident that if the endo angle constraints are “turned on”, i.e., if the dashed lines are treated as solid (locking

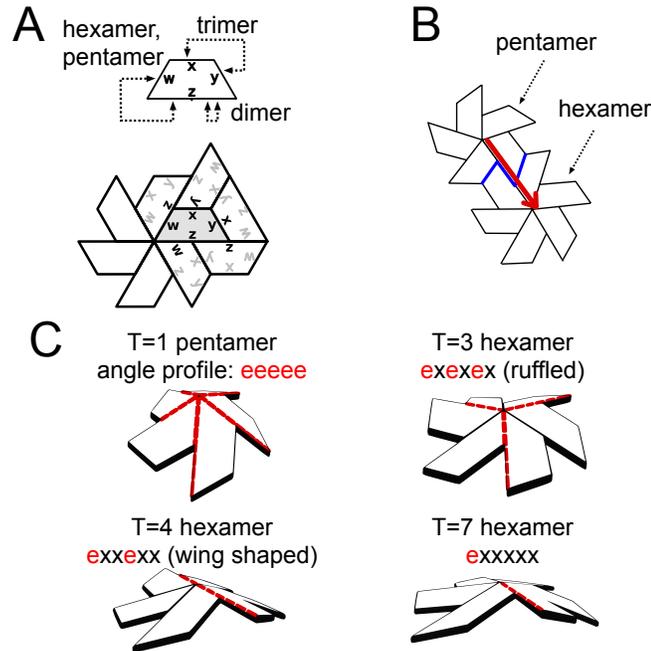
<sup>9</sup>From the various definitions of the canonical capsid of triangulation number  $T$  [1, 12], we have the number of hexamers per capsid equaling  $10(T - 1)$  and the number of pentamers per capsid equaling 12. Since we have 5 endo angles per pentamer, the average number of endo angles per hexamer that are imposed directly by pentamers must equal  $12 \times 5/[10(T - 1)]$  or  $6/(T - 1)$ .

certain dihedral angles at  $\sim 138.19^\circ$ ), only certain hexamers (possessing one or fewer endo angle constraints) will be allowed to sample at least two easily obtainable but distinct configurations. Specifically, those “flexible” hexamers must belong to  $T \geq 7$  capsids, where the average number of endo angles per hexamer ( $6/[T - 1]$ ), is less than or equal to 1. In this way, although endo angles do not directly constrain all angles in the hexamer (via the arrow depiction in Figure 4.3A), in some sizes ( $T = 3, 4$ ), all hexameric dihedral angles are effectively constrained due to specifically arranged endo angle constraints.

✓ All capsids may not be able to undergo buckling transitions.

To test this idea, we looked for the availability of accessible conformations to a capsid by physically perturbing (“squeezing and stretching”) dihedral angles within capsid *models* (deltahedra) of varying sizes ( $T$  numbers). The main assumption is that if the simplistic model is not able to sample alternative configurations, then the all-atom capsid that is constrained by simple geometry certainly will not. Here, the trimers are treated as rigid units (forming equilateral triangles, faces of the deltahedron). This is a reasonable assumption if subunit shapes are not greatly changed upon capsid buckling (as is noticed in the bacteriophage HK97, where the morphology change has little effect on the general shape of the subunit [64], while greatly modifying the hexamer pucker state [63]).

For each dihedral angle, we applied stretching and squeezing forces (that try to expand and contract the dihedral angles, discussed in the Materials and Methods section). The forces were incremented from 0 in small steps ( $0.00125\epsilon$  units, with cu-



**FIGURE 4.3: Simple geometry describes hexamer shape.** A canonical capsid subunit is shown in (A) with its bonding rules and one local environment. The pentamer-hexamer interface shown in blue in (B) possesses a curious effect where the hexameric angle adjacent to a pentamer must also be endo like (or pentameric) in nature. This effect—the endo angle constraint (shown as an arrow from the pentameric angle to that in the hexamer)—can be seen in natural canonical capsids as evidenced by the dihedral angle profiles in Figure 4.2A. The result, as shown in (C), is that hexamers belonging to different capsid sizes ( $T$  numbers) possess varying number of endo angles (red dashes) and may possess different hexamer shapes.

mulative forces ranging from 0 to  $\epsilon/8$ , where  $\epsilon$  is the bonded force constant of each bond/edge of the deltahedron), while minimizing the structure at every step. If there is no physical constraint geometrically placed upon the specific hexameric angle (on account of the architecture of the model), then the forces will cause a change in the structure and the recorded energy will remain at zero. If constrained, the capsid will be relatively unyielding to the forces, and the energy will increase harmonically with each step only to fall back into its original state after forces are lifted. Dihedral angle

tests showed that all hexameric dihedral angles within the  $T = 1, 3$  and 4 capsids are rigid/constrained within our force regime.

However, analysis of the  $T = 7$  capsid model—where the number of endo angles per hexamer is 1 (i.e., where  $6/[T - 1] \leq 1$ )—shows that some hexameric angles are able to sample an alternative conformation (indicated by the availability of multiple local minima and hexamer configurations in Figures 4.4A and B respectively). The change is not instantaneous upon application of infinitesimal force, but is dependent on overcoming a small energy barrier (akin to going through a transition state). Our results indicate that buckling transitions that require the sampling of two distinct conformations may be available only to  $T \geq 7$  capsids (as evidenced in our  $T = 7, 13$  models). However, we stress that not all large capsids may possess this ability even at a simple geometric level. For example, the  $T = 9$  capsid/deltahedron, which is purely icosahedral in shape (with 20 triangulated facets of 27 subunits), may not possess the ability to easily sample multiple configurations on account of its idealized icosahedral shape (which is purely convex and hence geometrically highly stable).

It is noteworthy that buckling of capsids represented by continuum elastic shells have been performed before, where interesting relationships between radius, capsid size, and sphericity were established [65, 66]; however, in these studies, the predictions made have yet to be applied to capsids of specific  $T$ -numbers. The continuum models neglect molecular/geometric features of the capsid (such as hexamer shape), and are therefore not analogous to our investigations, which are centered around subunit-shape-resolved models. It will be interesting to see if inferences/predictions from continuum and geometric methods converge.

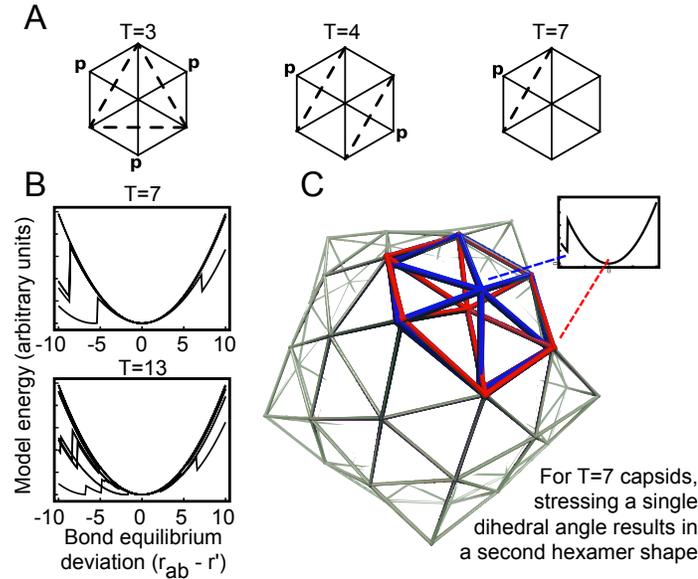


FIGURE 4.4: **Only  $T \geq 7$  capsid models appear to “buckle”.** Hexamer graphs taken from various capsid sizes (A; where subunits are represented as solid-edged triangles) show that, geometrically, only hexamers from  $T = 7$  canonical capsids (or larger) may undergo changes in shape while maintaining monohedrality (defined in Section 3.1.1.3 as maintaining subunit shape); this occurs in light of the endo angle constraints (shown effectively as dashed edges) imposed by pentamers (marked by P). Forces applied onto individual dihedral angles within capsid models (A; see Section 4.5.4) indicate that  $T < 7$  capsid models are geometrically rigid upon application of small forces on dihedral angles (indicated by parabolic force-energy profiles and singular minima, shown in Figure A.3), while the geometry of  $T \geq 7$  capsids appear to allow for specific dihedral angles to sample multiple values (shown here for  $T = 7, 13$ ). The result, especially for  $T = 7$  capsid models, is that hexamers within the capsid are able to sample two distinct configurations (blue and red hexamers), a result that parallels buckling transitions in theoretical [67] and experimental studies of the  $T = 7$  capsid [63, 64].

### 4.3.1 Buckling transitions versus other maturation events

We distinguish between what we call “true” buckling transitions and other maturation events such as capsid swelling (or its inverse: shrinking). Buckling transitions are those transitions that allow a shell to sample two morphologies—one being more “spherical” and the other being more “faceted” or icosahedral—without undergoing major changes in subunit-subunit bondedness and subunit shape [64]. This excludes the other kind of maturation events—swelling [68–70]—that is theoretically available to

any capsid regardless of size. Also, those maturation events requiring gross change in subunit shape (e.g., as seen in Flaviruses [71]) are not considered here.

Swelling is primarily caused by weakening of interfaces (via pH modulation, ion depletion, electrostatic screening, etc.), which causes a radial capsid swell (its converse, “shrinkage”, happens when subunit-subunit interactions are strengthened). These events often accompany the introduction/removal of holes between subunits (commonly found at within trimers), which can not be modeled by simple monohe-dral tilings/deltahedra (as holes must be considered as additional tiles). Examples of swelling and shrinkage are the  $T = 3$  and unnatural  $T = 1$  plant viruses such as sesbania mosaic virus (that undergo swelling) [68, 72] and  $T = 4$  semliki forest virus and  $T = 16$  herpes virus (that undergo shrinkage from a swollen precursor to a finally more icosahedral looking capsid) [69, 70], all of which display holes in their expanded or swollen forms. These kinds of swelling/shrinking transformations comprise radial motions that have been given previous theoretical consideration [22, 73] and were not considered here.

### 4.3.2 The need for auxiliary proteins

Here, we established that the pentamer imposes its endo dihedral angle properties onto *adjacent* hexameric dihedral angles (Figure 4.3B), thereby constraining shapes available to adjacent hexamers. This, along with well recorded quasi-equivalent mechanisms (“switches”) such as order-disorder transitions (reviewed in [14, 15]) are adequate in ensuring the existence of both pentamers and hexamers *adjacent* to pentamers in small ( $T \leq 7$ ) capsids.

However,  $T > 7$  capsids possess more than one hexamer species, where the secondary hexamer type is no longer in contact with any pentamer. Such hexamers may not be directly influenced by the geometric endo angle constraints (and adjacent quasiequivalent mechanisms), and therefore, we argue, may need other (auxiliary) constraints to secure the shape of the isolated hexamer. It is interesting that, so far, all  $T > 7$  spherical capsids have been experimentally found to require auxiliary proteins to form native structures (noted in [14]). It is also interesting that, during *model construction*, all  $T \leq 7$  capsids did not require any additional constraints to ensure *uniform* hexamer shapes, while the second hexamer that is isolated from the pentamers in the  $T = 13$  model was able to sample at least two distinct (and energetically viable) shapes within the capsid, resulting in a non symmetric and subunit-subunit bond-wise “complicated” capsid structure (“hexamer 2” of the  $T = 13$  capsid model in Figure 4.2B is an averaged version of positionally equivalent but architecturally varying hexamers). Based on experimental and geometric studies, we suggest that all  $T > 7$  capsids require auxiliary mechanisms (by means of proteins interaction, etc.) to maintain the shape of secondary hexamers.

Stating that  $T > 7$  capsids must need auxiliary proteins does not preclude the  $T \leq 7$  capsids from displaying auxiliary proteins—for any capsid size, auxiliary proteins may serve as an excellent mechanism for viral lifecycle control. Our statement only implies that  $T > 7$  capsids may be theoretically/geometrically excluded from forming all required capsomer shapes (to form the final capsid) without auxiliary help in the form of proteins or additional (currently unelucidated) mechanisms to assist in the formation of the secondary hexamers.

✓ Bigger ( $T > 7$ ) capsids need helpers to form correctly. Could these helpers be used as antiviral targets?

### 4.3.3 Auxiliary proteins vs. buckling availability

Some  $T > 7$  capsids are known to retain the auxiliary proteins within the final capsid (e.g., the  $T = 13$  birnavirus [49] and reovirus [74, 75]). This adds an interesting imposition onto  $T > 7$  capsids; for even if they theoretically could buckle, their present morphology may be “locked in” due to contact with the auxiliary proteins. If this is true, the presence of such auxiliary proteins may impede buckling of  $T > 7$  capsids, i.e., buckling transitions may be *practically* possible only for  $T = 7$  capsids. Currently, empirical data shows direct evidence of buckling transitions exclusively in  $T = 7$  capsids [63, 64, 76], supporting this hypothesis.

## 4.4 Concluding Remarks

How do capsids form different sizes? The theory of quasi-equivalence posits that the co-existence of the pentamer and hexamer allows for capsids of various sizes to exist [12]. Here, we showed, from empirical evidence and *ab initio* models, that *shapes* or puckers of the hexamers are strongly indicative of size within all available canonical capsids (opening the possibility of rational design of artificial nanoarrays and hexamer-shape-modifying drugs). After relating canonical capsids to geometrical entities—deltahedra—we were able to use such models and geometric concepts (e.g., endo angle constraints) to arrive at interesting (and empirically supported) general insights and predictions regarding modulation of capsid assembly (auxiliary protein requirements) and post-

assembly capsid transformations (availability of buckling transitions). Previous work on the capsid subunit [1] and current work on the entire capsid underline the usefulness of simplified but accurate geometric models in elucidating various capsid features, especially those of general import.

## 4.5 Materials and Methods

### 4.5.1 Natural capsids studied

We studied dihedral angles within x-ray structures of all natural capsids unambiguously denotable as canonical capsids (capsids that are representable by monohedral tilings) [1]. The stringency of these qualities is crucial to the dihedral angle comparisons, and so only a portion of those capsids deemed as “canonical” in [1] were studied (those with stricter adherence to monohedrality). The studied virus families,  $T$  numbers, and PDBIDs obtained from the capsid repository VIPERdb [46] are as follows: *Nodaviridae* ( $T = 3$ ): fhv (available only in VIPERdb [46]), 1nov, 2bbv, 1f8v; *Sobemoviridae* ( $T = 3$ ): 1smv, 1x35, 1f2n, 4sbv, 1ng0; *Tombusviridae* ( $T = 3$ ): 1opo, 1tnv, 1c8n, 2tbv; *Tetraviridae* ( $T = 4$ ): 1ohf; *Siphoviridae* ( $T = 7$ ): 2frp, 2ft1, 2fs3, 2fsy, 1ohg; *Birnaviridae* ( $T = 13$ ): 1wce.

### 4.5.2 Analysis of angles within natural capsids

Looking at dihedral angle similarities within two quasiequivalent interfaces (say, between adjacent subunit pairs A-B and C-D; an example of a pair “1”-“2” is shown in Figure 4.1A) becomes easy when dealing with capsids representable as monohedral

tilings (canonical capsids). This is because the subunits within such capsids have little subunit-subunit architectural variability (interface-controlling quasi-equivalent switches notwithstanding). Consequently, to check for the similarity between two interfaces A-B and C-D, one need only structurally align the carbon alpha traces of AB and CD (both treated as rigid units instead of two proteins) and calculate the normalized root mean square deviation (RMSD). Low RMSD values indicate that the dihedral angles between subunits A and B and subunits C and D are similar (or identical, if the RMSD is 0). From these analyses, we gathered  $T$ -specific dihedral angle patterns for hexamers (Figure 4.2A).

### 4.5.3 Creating capsid models

A majority of capsids possess trapezoidal subunits [1], whose interactions are described in Figure 4.3A. It is trivial to conclude that subunit trimers caused by “x-y” interactions will remain rigid as a unit if the subunits remain generally rigid. Consequently, it is acceptable to treat each coplanar trimer as a single face, i.e., monohedral capsid models of  $60T$  subunits may be represented as  $20T$  equilateral triangle faced polyhedra otherwise known as *deltahedra*. The simplest deltahedron, the  $T = 1$  deltahedron, is the twenty faced icosahedron. We created these deltahedra by creating duals of deltahedra in Euclidean space (which, are, interestingly models of buckyballs) and then obtaining the deltahedra from those duals.

We produced the deltahedral dual (buckyball) by first generating the graph connectivity using the spiral code method described by Fowler *et al.* [36]. From this abstract graph description, for each  $T$  number, we constructed a planar graph (Schlegel

diagram) of the abstract graph using an algorithm modeled around one described by Bor Plestenjak [77]. This planar graph was then wrapped around a sphere (using a non-linear plane to sphere projection). The final minimized structure (minimized so that all edge lengths are equal, will resemble an icosahedral buckyball. The buckyballs were transformed into their duals (whose graphs and general shape resemble the required deltahedra). This structure was then minimized to insure that deltahedra edges are equal (and set arbitrarily to 18Å). For the capsids studied ( $T = 1, 3, 4, 7, 13$ ), the final minimized structures were found to be in the lowest energy possible, and were mostly icosahedral (the  $T = 13$  capsid was the exception; see the note on auxiliary proteins in the Results and Discussion section). These structures were used for the final analysis of (1) hexameric dihedral angle configurations and (2) availability of buckling transitions.

#### 4.5.4 Assaying subunit-subunit dihedral angle constraints

For our capsid structure to possess multiple *interchangeable* configurations, one would expect a range of allowable values for at least some dihedrals within the capsid (especially within the hexagonal regions). We start with the obtained deltahedra and define the dihedral angles across any edge  $ij$  as  $\phi_{ij}$ . Each edge is shared by two equilateral triangles (shown in isolation from the entire deltahedron in Figure 4.5). The relationship between the dihedral angle  $\phi_{ij}$  and the distance between the non-common point  $r_{ab}$  is

$$\sin(\phi_{ij}/2) = r_{ab}/2m$$

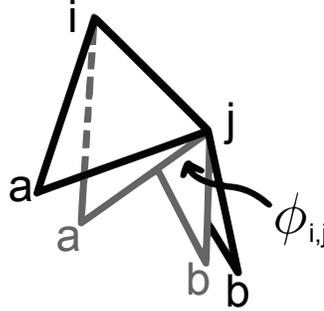


FIGURE 4.5: **Squeezing and stretching dihedral angles.** The dihedral angle  $\phi_{i,j}$  between two equilateral faces sharing edge  $\{i,j\}$  (shown in two configurations) is dependent on the distance between  $a$  and  $b$  ( $r_{ab}$ ).

where  $m$  is the height of the equilateral triangles. Therefore, adding a single dihedral restraint across the edge  $\{i,j\}$  is analogous to adding a new bond to the system with potential energy

$$E_c = \frac{1}{2}k_c(r_{ab} - r'_{ab})^2$$

It is imperative that the force constant  $k_c \ll k_{rest}$ , where  $k_{rest}$  is the strength of the bonds making up the deltahedron. This is required as we are studying elastic deformations of dihedral angles (and the dihedral bond) and not the equilateral faces of the deltahedra (although small deviations in shape are acceptable).

We assayed the effect of applying stress onto a dihedral angle with respect to resulting energy change. The study is performed using the following algorithm:

**Initialization step:** (a) Identify the edge ( $\{i,j\}$ ) whose dihedral angle is to be studied. (b) Assign a restraint energy term  $E_c$  as shown above to the appropriate atom pair ( $\{a,b\}$ , in the Figure 4.5). (c) Assign  $r'_{ab} = r_{ab}$ , where  $r_{ab}$  is the length between atom pair  $\{a,b\}$  in initially obtained (embedded) deltahedron. This ensures that at the first step all energy terms equal zero (since the deltahedron is minimized).

---

**Cycle (till  $|r'_{ab}| \leq |r'_{max}|$ ):** **(a)** Assign  $r'_{ab} = r'_{ab} + step\_size$ . This will cause a force to be applied onto  $a$  and  $b$  as the  $\{a, b\}$  bond length will not be at its equilibrium value. **(b)** Allow the structure to relax by energy minimization. At this point, we obtain the *total* energy of the new structure.

Next chapter ...



Evolution

Simple designs rule

Nature secretly discriminates

A colorful periodic table explains much

## Chapter 5

# Evolution of the virus capsid

Spherical viruses are highly efficient infecting machines, a fact that was made quite clear to the free-food-eating journal club dwellers at Scripps, where a majority of the graduate students present were struck by explosive diarrhea (a casual side effect of the the norovirus life cycle, a spherical virus of gastrointestinal proclivity). Other spherical viruses affect humans in other odious ways, causing suffering by way of deathly illnesses and crop destruction. How did the viruses we see today come to be such efficient animals? This question is primarily an evolutionary one, and this chapter focuses on asking whether capsid evolution and natural selection are mathematically describable.

Even though evolution plays an unambiguously important role in the working of the virus, so far, general evolutionary pressures shaping capsid design have remained elusive. Here, we present a periodic table of virus capsids based on geometric principles, which uncovers a strong, overarching and unprecedented evolutionary pressure based on the notion of reducing the capsid's design complexity (or "hexamer complex-

ity”,  $C^h$ ). This periodic table also offers geometric explanations to other capsid properties (rigidity, pleomorphy, auxiliary requirements, etc.) in the context of a systematic framework not presented before. Available virus structure databases and other published data reiterate the predicted geometry-derived rules, reinforcing the role of geometry in the natural selection and design of virus capsids.

This chapter is modified from the following journal manuscript: R.V. Mannige and C.L. Brooks III. (2010) Periodic table of virus capsids: implications for natural selection and design. *Public Library of Science ONE*, in press<sup>1</sup>.

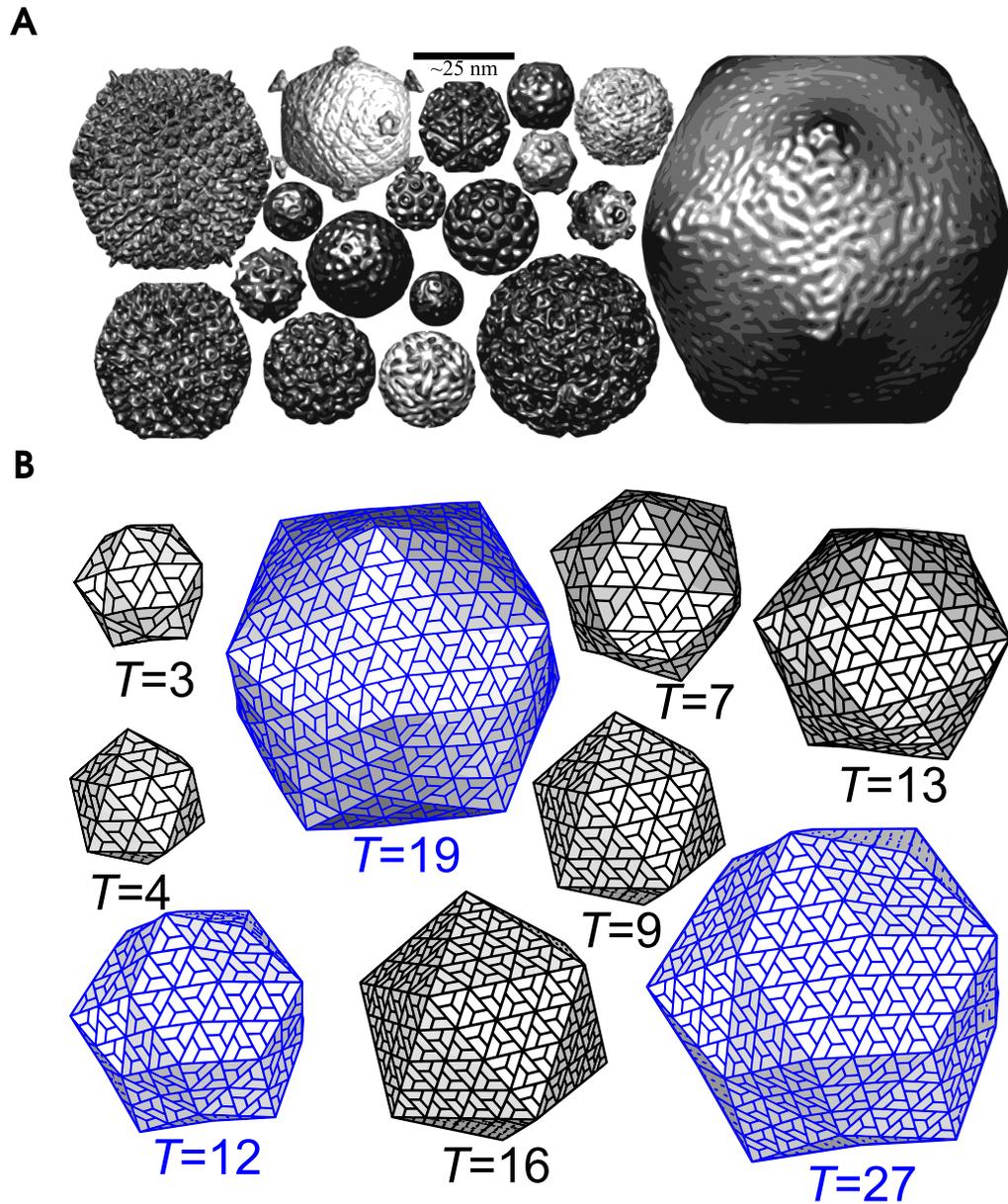
## 5.1 The mystery of the missing capsids

Half a century of empirical data (and Chapter 4) tells us that a large array of capsid sizes exist that range from tens to many thousands in subunit composition (see Figure 5.1A) [8]. Still, some sizes are rarer than others (the blue ones in Figure 5.1B<sup>2</sup>), an observation that puzzled structural virologists as early as 1961 [11, 12]. The cause for this discrepancy remains unexplained. Why are some capsid sizes not seen even today? Are specific spherical viruses disadvantaged from an evolutionary perspective? Or have we just not looked enough or in the right places? In this Chapter, we present a conceptual framework useful in providing answers to these questions, while arriving at interesting observations about capsid classes, distributions, morphologies and mechanical properties. The following points will be very useful in “solving” this rather interesting mystery:

---

<sup>1</sup>For a copy of this manuscript, please contact me at [ranjan@umich.edu](mailto:ranjan@umich.edu)

<sup>2</sup>Yes, they are blue.



**FIGURE 5.1: Inconspicuously missing capsids.** The interesting variety of capsids seen in nature (A) belies a sneaky evolutionary occult, since there are various theoretically sound sizes—e.g., the blue ones in (B)—that are either rare or not found yet in nature. Why?

1. Trapezoidal subunits can assemble, theoretically, into capsids of any triangulation ( $T$ ) number (or size), i.e., the capsid design is scalable (Section 2.1.3).

2. Endo angles emanate from pentamers and constrain *specific* dihedral angles within capsids (Section 4.2.1).

3. The pattern of endo angles in the capsid is size specific, but not  $T$ -specific (it is dependent on  $h$  and  $k$  instead, where  $T$  is a function of  $h$  and  $k$ ; see Equation 2.1)

## 5.2 Revising old concepts

Spherical capsids of all observed sizes may be characterized by two integers,  $h$  and  $k$  (first discussed by Goldberg [13]; see Section 2.1.3), which describe the number of hexamers ( $h+k-1$ ) one would have to “walk over” to get from one pentamer to an adjacent pentamer within a completed capsid (the walk is shown as rays in Figure 5.2A) [12]. As a rule, a longer “walk” indicates the presence of more hexamers in the structure (while the number of pentamers will always be the same at 12), which means a larger capsid. A useful metric for capsid size—the *triangulation number*,  $T^3$  (discussed in Section 2.1.5)—was introduced by Caspar and Klug [12], where a capsid of triangulation number  $T$  is expected to be comprised of  $60T$  subunits, or 12 pentamers and  $10(T-1)$  hexamers, i.e.,  $T$  is a quantitative metric for capsid size. We now show, using “endo angles”, that  $h$  and  $k$  (and not  $T$ ) are sufficient in providing a useful capsid classification schematic.

---

<sup>3</sup>Where  $T = h^2 + hk + k^2$  and  $h, k$  are the size indicating numbers.

- ✓ Next, we will use the concept of the *endo angle constraint* to draw connections between a capsid classification scheme (developed below) and hexamer shapes present within a capsid.
- ✓ These concepts will then allow us to arrive at a metric for capsid complexity (hexamer complexity), which is useful in explaining and predicting various structural and evolutionary properties of the capsid.

## 5.3 Endo angles, classification, and history

### 5.3.1 Endo angle propagation and termination rules

As described discussed in Chapter 4 and a publication [2], endo angles (specific subunit-subunit dihedral angles) that originate from the pentamer must propagate through the adjacent hexameric lattice in what we call *endo angle propagation* (it is a constraint imposed by pentamers onto adjacent hexamers, which is depicted as arrows in Figure 5.2B). This constraint has been useful in predicting the existence of various distinct hexamer shapes [2]; here, hexamer shape is defined by the hexamer pucker or subunit-subunit planar angles within the hexamer (The number of hexamer shapes available are enumerated in Section B.3). Figure 5.2C-D shows how multiple endo angles within the *capsid* (not just hexamer) must interact, which is a new development.

### 5.3.2 Endo angles classify capsids

We find that the endo angle patterns produced by interacting endo angles within the capsid (Figure 5.2) ensure the emergence of three general morphological classes (Fig-

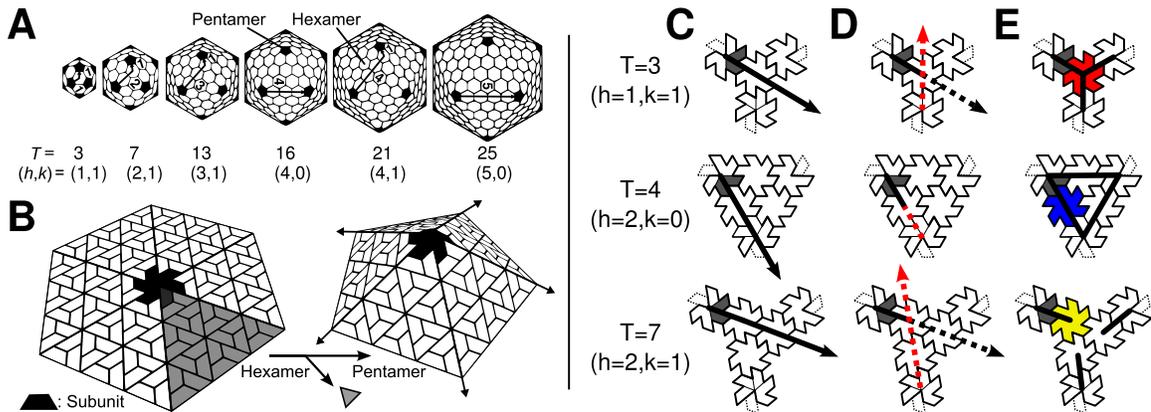


FIGURE 5.2: Capsids are scalable (A) and are formed of 12 pentamers (represented as darkened pentagons) and a variable number of hexamers. The act of producing a pentamer from a hexamer (B) imposes pentameric dihedral angle values (“endo angles”) onto its neighboring hexameric angles (discussed in Section 4.2.1 and Ref. [2]) which propagate through the hexamers (depicted by arrows in B) in what we call endo angle propagation. We define *endo angle termination rules* (C-E) for the three smallest capsids possessing hexamers ( $T = 3, 4, 7$ ) within a “face” (a triangular facet containing hexamers and three adjacent/neighbor pentamers). An endo angle (black ray) propagating from the shaded subunit-subunit interface belonging to a pentamer (C) is challenged and terminated by another endo angle (D, red dotted ray) propagated from a neighboring pentamer, not completely visible for  $T = 4$ ), resulting in hexamer shapes and capsid endo angle features (E) that are  $h$  and  $k$  specific. In particular the differences in  $h$ - $k$  relationships ensure hexamers of distinct shapes per capsid size (distinctly colored). This will be useful in arriving at a classification scheme.

ure 5.3) differentiated by their  $h$ - $k$  relationship: class 1 (described by the relationship  $h > k = 0$ ), class 2 ( $h > k > 0$ ), and class 3 ( $h = k$ )<sup>4</sup>.

### 5.3.3 Early attempts at size “existence theorems”

When this classification system crystallized, I was quite happy, and immediately set out to answer the following question: are the classes telling us something about capsid abundances and therefore virus evolution? That, it seems, was a premature question, since the same classification system was recognized qualitatively by not one but two

<sup>4</sup>Henceforth, we assume that  $h \geq k$  for simplicity’s sake, since, for our discussions, the difference between chiral  $l$  and  $d$  class 2 capsids is inconsequential. This is the case because the inherent connectivity describing both chiral forms are identical, and so physical and geometric properties (such as shape and rigidity) of the two chiral forms will be identical.

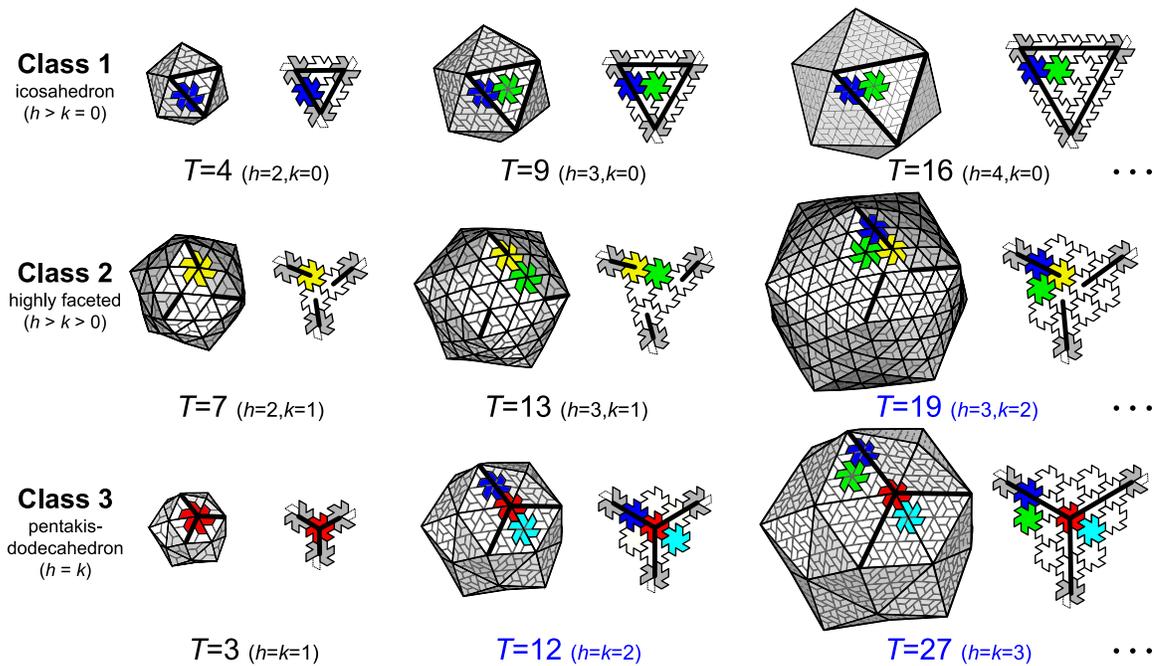


FIGURE 5.3: The three virus capsid classes. All canonical capsids (made up of trapezoidal subunits) may be built from a single type of pentamer and a repertoire of distinct hexamer shapes (colored distinctly only once in each capsid; also described in Figure B.1). The hexamer shape is described by the number of endo angles it displays. Endo angles are depicted as bold lines within a “face” in its isolated (right) and capsid environment (left) for the first three capsid sizes in each class (excepting  $T = 1$ ).

groups [11, 12] in the early 1960s!<sup>5</sup> To add to my disappointment, they had used the classification for purposes identical to mine: to explain absences in the capsid size diversity or  $T$ -range. Specifically, class 2 capsids (in our schematic) had not yet been observed, and both reports postulated that capsids from this class must be absent for specific (but distinct) physical reasonings [11, 12].

Since then, capsids from all three classes have amply been seen (abundances are reported in Table B.1), i.e., the classification system can not be used to make direct predictions about capsid existence. Consequently, this topic, which we are readdressing now, appears to have been latent since 1962.

<sup>5</sup>Fig. 10 in Ref. [11] and Fig. 8 in Ref. [12]; although both neither account directly linked  $h$  and  $k$  to class type.

## 5.4 Introducing Hexamer complexity ( $C^h$ ).

The utility of the class system is not entirely lost, however; Figure 5.3 allows us to obtain the minimum number of distinct hexamer *shapes* (discussed in Ref. [2] and defined in Figure 5.2) required to form a canonical capsid of specific capsid size. We introduce the *hexamer complexity* ( $C^h$ ) as the number of distinct hexamer shapes present in a capsid (a higher number of distinct hexamer shapes per capsid reflects a higher  $C^h$ ). One may obtain  $C^h$  by counting the number of distinctly colored (shaped) hexamers in Figure 5.3 (alternatively, an equation to calculate  $C^h$  from  $h$  and  $k$  is given in Equation B.18)<sup>6</sup>. We reason that capsids with higher  $C^h$  are evolutionarily disfavored.

### 5.4.1 High $C^h$ capsids require more auxiliary control during formation.

Evidence indicates that capsid formation is nucleated [78], often starting with a single capsomer species (e.g., pentamers [79])<sup>7</sup>, and proceeding to completion by the addition of small subunit clusters (or single subunits). In  $T = 1$  capsids, where all subunits are in identical/equivalent environments [12], nucleated assembly will be possible with no additional machinery (except for the predefined angle of incidence for each subunit-subunit interaction site). However, the formation of two or more capsomers from a single interaction site will require the employment of additional machinery to ensure high yields of the native state. For example, quasi-equivalent switches [14, 15] are required for the proper assembly of capsids containing two distinct capsomers—a pen-

<sup>6</sup>Note that in our schematic, hexamers in distinct environments are allowed to possess the same shape.

<sup>7</sup>For the purposes of this Chapter, a capsomer is a generally symmetric cluster of either five or six subunits that are the sole building blocks of a capsid.

tamer and one type of hexamer (i.e.  $C^h = 1$ ). The addition of a second hexamer shape ( $C^h = 2$ ) necessitates the requirement of a second mechanism such as auxiliary proteins [80] for proper assembly (discussed earlier in theory [2] and evidenced from the observation that all recorded  $C^h > 1$  or  $T > 7$  capsids are known to require auxiliary proteins for assembly [14]).

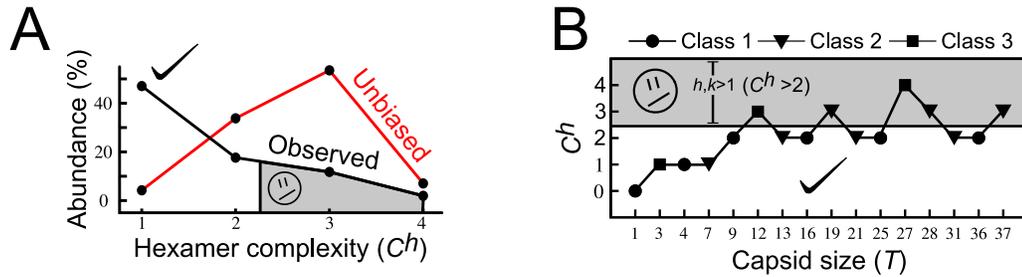


FIGURE 5.4: Evolutionary discrimination of spherical capsids. **(A)** As predicted by the inverse  $C^h$  rule, capsids with high hexamer complexity are under-represented in nature as evident in the observed versus unbiased capsid abundances (% of families that display capsids of specific  $C^h$ ). **(B)** Although  $C^h$  is directly related to  $(h,k)$ , it is not conveniently correlated with capsid size ( $T$ ) or class (symbols), which could explain why capsid size discrimination remained elusive until now.

### 5.4.2 Capsid $C^h \propto 1/\text{capsid abundance}$

For virus capsids requiring more distinct hexamer shapes (larger  $C^h$ ) additional mechanisms to stabilize those new shapes at exactly the right positions within the forming capsid are likely also needed<sup>8</sup>, the interplay of which, we propose, would be theoretically possible to choreograph but unduly complex. Accordingly, we predict that canonical capsids with larger  $C^h$  will be encountered with a lower frequency in nature<sup>9</sup>.

<sup>8</sup>Lest off-pathway and fatal configurations would dominantly form.

<sup>9</sup>It is beyond any doubt that complexity is often not the sole criterion for natural selection. In fact, if that was the case then humans would never be given the chance to come into existence. But alongside natural selection arises the notion of the niche, that states that, among organisms that live within a niche and that compete for the same natural resources, the most efficient design will likely prevail. This comes into play when we consider spherical viruses that are dissimilar in  $C^h$  but operate under identical host and reproductive constraints. In those situations, the capsid with a simpler and more efficient design,

Support for this relationship (that higher  $C^h$  will be encountered with lower frequency in nature) is presented in Figure 5.4A (and discussed further in Figure B.2), where there is an inverse correlation between capsid  $C^h$  (calculated using Equation B.18) and observed capsid abundance (for  $C^h > 0$ )<sup>10</sup>. However, this is not the case for unbiased capsid distributions (red line) where we assume no evolutionary favoritism (i.e., if we assume that each capsid size or  $T$  is equally probable to exist for the size range observed;  $T = 1$  through 219). Also apparent in this data is the observation that  $C^h > 2$  capsids are under-represented by a factor of  $\sim 12$  ( $\sim 63\% : 5\%$  for unbiased vs. observed capsid abundances) when compared to the calculated distributions for the observed size range<sup>11</sup>. This suggests that a large evolutionary pressure in aversion to high hexamer complexity may be at play in nature.

### 5.4.3 Capsid $C^h$ is related to class $(h,k)$ not size ( $T$ ).

Although not directly relatable to capsid size ( $T$ ) and class (Figure 5.4B),  $C^h$  is easily obtained from the Goldberg parameters  $h$  and  $k$  (Equation B.18) from which we can show that  $C^h > 2$  when both  $h > 1$  and  $k > 1$  (Table B.2).  $C^h$  rules are concisely reiterated in periodic form in Figure 5.5C such that, through each period (row), hexamer complexity ( $C^h$ ), class number, and triangulation number ( $T$ ) increase from left

to right, allowing us to predict that capsids belonging to the right side of this table i.e., those with low  $C^h$ , will be more efficient than the higher  $C^h$  capsid in assembling, and therefore propagating.

<sup>10</sup>Data, listed in Section B.11, were pooled from EM and X-ray structure repositories [8, 81]. We did not distinguish between capsids containing external lipid membranes and those that do not, since, often, such lipids are post assembly features (e.g., herpesvirus [82]).

<sup>11</sup>If we calculate expected distributions for a more conservative range of  $T = 1$  through 31, the unbiased value is still  $\sim 6$  times higher than our observed 5% at  $\sim 29\%$ .

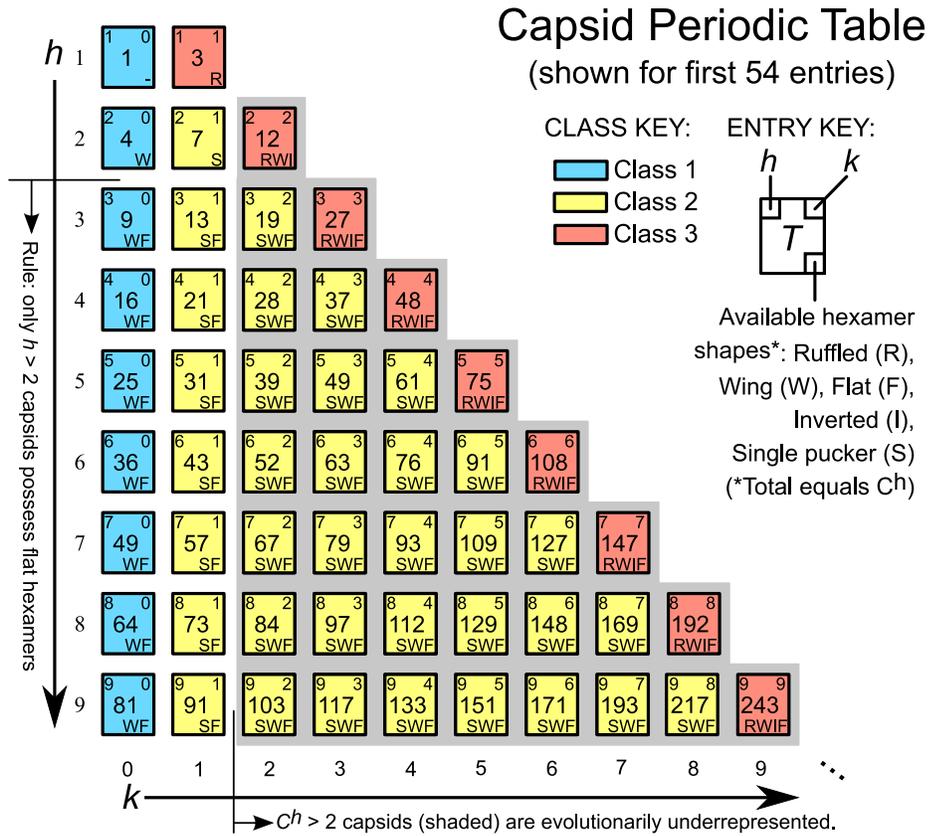


FIGURE 5.5: Periodic table of spherical capsids. Trends in  $C^h$  are easily discerned from the periodic table, where, in each period (row),  $T$ , class number and  $C^h$  increase (or remain the same), while trends in other capsid properties such as rigidity may also be deciphered.

$(h, k > 1)$  are evolutionarily disfavored<sup>12</sup>. Since capsid class describes distinct geometries, we expect that this table will also be useful in describing physical properties such as capsid rigidity.

Our complexity rules, although arising from geometric analysis of *canonical* capsid models [2] (further discussed in Section B.1), appear to be applicable to almost all observed capsids, indicating that hexamer complexity may be a universally important

<sup>12</sup>Note that there is no one-to-one mapping of  $T$  on  $(h, k)$ ; e.g.,  $T = 49$  may be constructed from  $(h, k)$  pairs  $(7, 0)$  and  $(5, 3)$  assigned to classes 1 and 2 respectively, i.e., some  $T$  numbers will be repeated in the periodic table.

concept (if we include only canonical capsids [1], the number of  $C^h > 2$  capsids reduce to zero!). We will shortly discuss the few “rule breakers”.

#### **5.4.4 Evolution/natural selection vs. design**

At this point, it is important to distinguish design from evolution. From a design perspective, capsids of any size (or  $T$  number) may be easily “built” from an intricate set of rules, like in a Lego<sup>®</sup> construction kit. However, we suggest that, from an evolutionary perspective, the probability of “existence” is contingent upon whether a capsid structure can be produced via easily manageable assembly mechanisms. This is especially interesting since capsids with high  $C^h$  do not indicate larger size but just a more complicated design. E.g.,  $T = 12$  capsids, although smaller than  $T = 13$  and  $T = 16$  capsids, are vastly more complicated and under-represented in nature. Although our complexity-based rules imply a form of evolutionary pressure, other pressures will likely exist, whose effects might be overlaid to give a more intricate understanding of the available capsid distributions (e.g., geometrically simple  $T = 1$  capsids, although low in  $C^h$ , may be selected against due to restrictions on genome size; see Figure B.2).

### **5.5 Understanding the rule breakers and charting a phase diagram**

#### **5.5.1 Rule breakers**

There are two major groups of  $C^h > 2$  outliers/rule-breakers—the small ( $T < 31$ ) and large ( $T > 100$ ) group—that display distinct characteristics. Markedly, most of the

small rule-breakers possess an internal support/core of lipid or protein [83–85], or display unusually high number of protrusions and putative proteins associated with their capsomers [86]. These examples indicate that evolutionary constraints of a geometric nature placed upon isolated capsids may be overcome by employing “universal scaffolds” such as protein/bilayer cores and excessive auxiliary proteins useful in maintaining all distinct capsomer/hexamer shapes (Recently, another small rulebreaker not used in our study was also shown to have an internal membrane [87]). We predict that, generally, the amount of “extra subunit density” in the electron density of a capsid is inversely related to  $C^h$ .

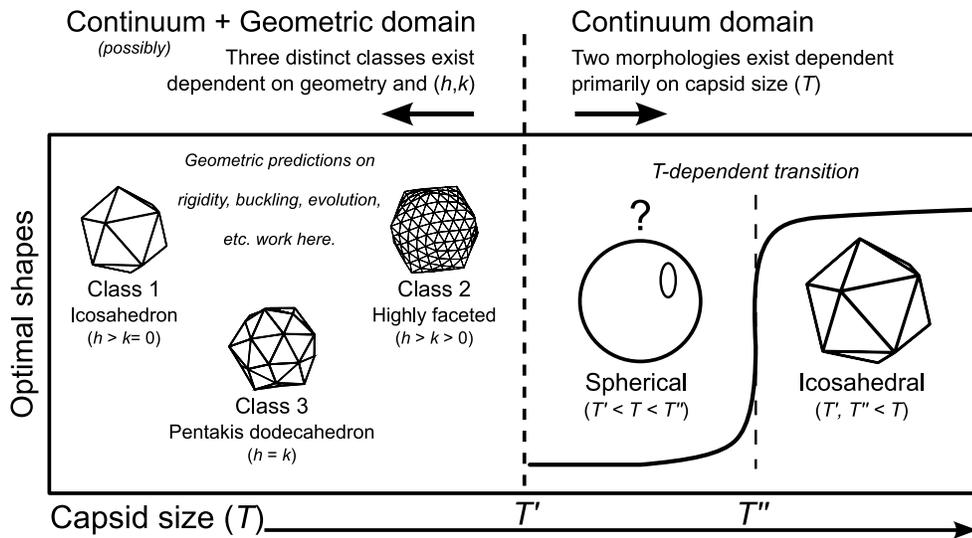
The remaining three (big) capsids [88, 89] that break our geometric rules possess thousands of subunits. This is interesting, since these capsids are possibly of large enough size that the “discreteness” (or geometric/molecular subunit nature) of the capsid shell is less influential than the shell’s bulk properties [90], which would allow for those capsids to be exclusively modeled as elastic shells. This knowledge is helpful in constructing a proposed phase diagram for spherical capsids (Figure 5.6).

✓ The rule breakers (high hexamer/design complexity structures) are either very big ( $T > 100$ ) or are small with permanent scaffolds.

### 5.5.2 Phase diagrams

As described above, it is inevitable that, at a certain triangulation number ( $T'$  in Figure 5.6), the capsid morphology will not be influenced by molecular/subunit/hexamer properties (where geometric relationships hold), beyond which capsids may be modeled exclusively by continuum elasticity theory [90]. Work using continuum elasticity has shown that only two capsid shapes must exist—spherical and icosahedral, and that

the transition between them is demarcated by the capsid's *Föppl-von Karman number* [65, 66], which is directly proportional to  $T$  (especially if the size of the subunit is generally the same)<sup>13</sup>. It is then interesting that the large ( $T > 100$ ) capsids are all icosahedral in shape, no matter what class they are present in. In our “phase diagram”, we also introduce a theoretical capsid size  $T''$  (Figure 5.6 arbitrarily assumes that  $T'' > T'$ ) that differentiates between the sphere-icosahedron boundary predicted by continuum elasticity theory (the sigmoidal curve in Figure 5.6 denotes the change in sphericity discussed before [65] that is dependent on  $f$  and hence capsid size,  $T$ ).



**FIGURE 5.6: Spherical capsid phase diagram.** We describe two specific capsid sizes that remain to be elucidated ( $T'$  and  $T''$ ; the diagram arbitrarily assumes that  $T'' > T'$ ).  $T'$  describes the limit of the geometric domain, beyond which our geometric assumptions and predictions may not hold. We expect that all capsid sizes greater than  $T'$  will be exclusively described by continuum elasticity. We also expect that, beyond  $T'$  (i.e., in the purely continuum domain), the *Föppl-von Karman number* ( $f$ ) [65, 66] that dictates spherical vs. icosahedral morphology will depend primarily on  $T$ , and so there will be a capsid size ( $T''$ ) that demarcates the allowance for spherical and icosahedral morphologies in the purely continuum regime (the sigmoidal curve represents the dependence of  $f$  and hence morphology on  $T$ ). These assumptions consolidate all observed instances of spherical capsid morphology.

<sup>13</sup>In the continuum domain: The Föppl-von Karman number is  $f = YR^2/\kappa$  [65], where  $Y$  and  $\kappa$  are bulk properties of the protein subunit, and  $R$  is the capsid's radius. If we assume that proteins, at an approximation, have similar size and bulk properties, then  $f$  will be directly proportional to  $R$  and therefore also to  $T$ .

The phase diagram brings to light a curious absence. So far, “hexamer complexity” was used to explain the elusiveness of certain capsid sizes (the  $h, k > 1$  capsids peppered through size or  $T$ -space). There is, however, a swath of the  $T$ -space (so far, between  $T = 31$  and 147) where no capsids, to our knowledge, have been reported. Beyond this  $T$  number only purely icosahedron-shaped capsids have been observed. It will be interesting to see whether capsids from this region ( $31 < T < 147$ ) will be found in the future, and if so, what their shapes will be. Note that the diagram, although fitting all observed data, represents *one* situation where  $T'' > T'$  which does not need to be true (since we could also have  $T'' \leq T'$ , where “spherical capsids” in the continuum domain will never exist).

### 5.5.3 Continuum theory and the phase diagram.

In continuum elasticity theory,  $f$  describes capsid morphology which ranges from completely spherical (for smaller  $f$ ) to completely faceted or icosahedral (for larger  $f$ ). In the geometric sense, the sphericity of capsids in the class system decrease in the following manner: class 2 > class 3 > class 1. Within the geometric domain ( $0 < T < T'$ ), it is certain that shape is dictated by capsid class (described by  $h-k$ ) and not directly by capsid size (for example,  $T = 16$  capsids are more faceted than  $T = 19$  and  $T = 21$  capsids; and  $T = 25$  capsids are more faceted than  $T = 28$  and  $T = 31$  capsids). In light of this, if the continuum domain ranges to even the smallest capsids, we predict that  $f$  would increase non-monotonically (i.e.  $f$  would fluctuate) through capsid size ( $T$ -space) till  $T'$ , after which it will increase relatively smoothly and monotonically (w.r.t.  $T$ ) due to the absence of geometric (or  $h-k$  based) influences (Figure 5.6). It will be

interesting to find whether theoretical calculations are able to reiterate this trend, as it would then be possible to obtain an estimate for  $T'$ .

## 5.6 Further implications

### 5.6.1 Classes, shapes and buckling

Because capsids from different classes display markedly different geometries, they are bound to display different physical properties. For example, since icosahedra and pentakis dodecahedra are geometrically rigid<sup>14</sup>, class 1 and class 3 capsids that employ such shapes should be unable to undergo buckling transitions (crucial virus life cycle events) [63, 91]. However, we expect class 2 capsids to be able to undergo such transitions due to their highly faceted (“harmonica like”) geometry, which allows for comfortable sampling of alternative structures. Also, class 1 and 3 capsids display a complete cage of endo angles spanning from pentamer to adjacent pentamer that serves as a frame to rigidify the structure. This is not the case for class 2 capsids, where endo propagations are prematurely terminated. Experimental work on one class 2 capsid, HK97 [63, 91], along with studies on capsid models ( $T = 1$  through 7 and 13) [2] lend credence to this hypothesis. Still, the existence of naturally buckling capsids of sizes other than  $T = 7$  remains elusive.

---

<sup>14</sup>This is a basic result of geometry.

### 5.6.2 *T*-switching and pleomorphy

The periodic nature of capsid hexamer content (Figure 5.5C) is also useful in understanding “*T*-switching”: a process that permits *canonical* capsid subunits to more easily sample capsid architectures containing similar hexamer shapes. This was shown to be true for a  $T = 4$  capsid subunit that, upon mutation, exclusively formed a range of class 1 capsids [55] that have similar hexamer shapes. This allows for a segue to understanding currently intractable and deadly pleomorphic viruses like ebola and arenaviruses. For example, from the above *T*-switching rule, the available diversity of an arenavirus (described by the observation of  $T = 3, 4, 9, 12$  and 16 capsids in a single sample) [92] may only be explained if we assume that the biologically relevant form of the arena virus is the  $T = 12$  capsid (since it exclusively displays all hexamer species required for all the other listed capsid sizes excluding the flat hexamer, which allows us to assume that all other sizes are residual byproducts of inefficient  $T = 12$  capsid assembly). Other predictions of this sort are easy to compile from Figure 5.5C and remain to be completely developed, explored and validated.

### 5.6.3 Non-icosahedral capsids

Although the framework presented doesn’t appear to readily explain non-spherical capsids (some are just “slightly” non-icosahedral, such as the natively prolate phi29 capsids [51], while others are wildly different in form, such as ebola with its natively filamentous shape), those capsids, like their icosahedral counterparts, also display capsomer sub-structures (for example phi29 capsids contain pentamers and hexamers,

while there is evidence that filamentous ebola capsids may contain hexamers *as well as* octamers [93]). In light of this, the geometric constraints analogous to endo angles that affect capsomer shape may be useful in obtaining insights into non-icosahedral capsid morphology, behavior, and classification. It will be exciting to see whether incorporating the non-icosahedral capsids into an expanded capsid periodic table will be possible.

## 5.7 Take home

Hexamer complexity ( $C^h$ ) and the periodic table provide a framework that explains elusive evolutionary pressures on capsid design,  $T$ -switching, mechanics (rigidity/maturation) and pleomorphy. We anticipate that many other features may be overlaid upon the schematic developed here, allowing for a comprehensive and systematic understanding of, first, spherical capsids and then virus capsids of varied geometries.

# Bibliography

- [1] R.V. Mannige and C.L. Brooks III. Tiling nature of virus capsids and the role of topological constraints in natural capsid design. *Phys. Rev. E*, 77(5):051902, 2008.
- [2] R.V. Mannige and C.L. Brooks III. Geometric considerations in virus capsid size specificity, auxiliary requirements, and buckling. *Proc. Natl. Acad. Sci. USA.*, 106(21):8531–8536, 2009.
- [3] R.V. Mannige and C.L. Brooks III. Periodic table of virus capsids: implications for natural selection and design. *Public Library of Science ONE*, (in press), 2009.
- [4] N. Barquet and P. Domingo. Smallpox: the triumph over the most terrible of the ministers of death. *Ann Intern Med.*, 127(8 Pt 1):635–642, 1997.
- [5] G. Rice, L. Tang, K. Stedman, F. Roberto, J. Spuhler, E. Gillitzer, J.E. Johnson, T. Douglas, and M. Young. The structure of a thermophilic archaeal virus shows a double-stranded dna viral capsid type that spans all domains of life. *Trends Biotechnol.*, 101(20):7716–7720, 2004.
- [6] R.W. Hendrix. Hot new virus, deep connections. *Proc. Natl. Acad. Sci. USA*, 101(20):7495–7496, 2004.
- [7] M. Dyal-Smith, S. L. Tang, and C. Bath. Haloarchaeal viruses: how diverse are they? *Res Microbiol.*, 154(4):309–313, 2003.
- [8] M. Carrillo-Tripp, C.M. Shepherd, I.A. Borelli, S. Venkataraman, G. Lander, P. Natarajan, J.E. Johnson, C.L. Brooks III, and V.S. Reddy. Viperdb2: an enhanced and web api enabled relational database for structural virology. *Nucleic Acids Res.*, 37:D436–D442, 2009.
- [9] F.H. Crick and J.D. Watson. Structure of small viruses. *Nature.*, 177(4506):473–475, 1956.
- [10] D.L. Caspar. Structure of bushy stunt virus. *Nature*, 177:475–476, 1956.

- 
- [11] R.W. Horne and P. Wildy. Symmetry in virus architecture. *Virology.*, 28(15):348–373, 1961.
- [12] D. L. D Caspar and A. Klug. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp.*, 27:1–24, 1962.
- [13] M.A. Goldberg. Class of multi-symmetric polyhedra. *Tôhoku Math. J.*, 43:104–108, 1937.
- [14] J. E. Johnson and J. A. Speir. Quasi-equivalent viruses: a paradigm for protein assemblies. *J. Mol. Biol.*, 269(5):665–675, 1997.
- [15] J. Tang, J. M. Johnson, K. A. Dryden, M. J. Young, A. Zlotnick, and J. E. Johnson. The role of subunit hinges and molecular “switches” in the control of viral capsid polymorphism. *J. Struct. Biol.*, 154(1):59–67, 2006.
- [16] A. S. Fanning and J. M. Anderson. Protein-protein interactions: PdZ domain networks. *Curr Biol*, 6(11):1385–1388, Nov 1996.
- [17] S. C. Harrison. Peptide-surface association: the case of pdz and ptb domains. *Cell*, 86(3):341–343, Aug 1996.
- [18] K.E. Drexler. Molecular engineering: An approach to the development of general capabilities for molecular manipulation. *Proc. Natl. Acad. Sci. USA*, 78(9):5275–5278, 1981.
- [19] P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure*, 14(3):437–449, 2006.
- [20] A. Arkhipov, P.L. Freddolino, and K. Schulten. Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure.*, 14(12):1767–1777, 2006.
- [21] R. Zandi, D. Reguera, R. F. Bruinsma, W. M. Gelbart, and J. Rudnick. Origin of icosahedral symmetry in viruses. *Proc. Natl. Acad. Sci. USA*, 101(44):15556–15560, 2004.
- [22] R. Zandi and D. Reguera. Mechanical properties of viral capsids. *Phys. Rev. E*, 72(2):021917, 2005.
- [23] T. Chen, Z. Zhang, and S. C. Glotzer. A precise packing sequence for self-assembled convex structures. *Proc. Natl. Acad. Sci. USA*, 104(3):717–722, 2007.

- 
- [24] K. Van Workum and J. F. Douglas. Symmetry, equivalence, and molecular self-assembly. *Phys. Rev. E*, 73:031502, 2006.
- [25] D. C. Rapaport. Self-assembly of polyhedral shells: A molecular dynamics study. *Phys. Rev. E*, 70(5):1539–1555, 2004.
- [26] H. D. Nguyen, V. S. Reddy, and C. L. Brooks III. Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids. *Nano. Lett.*, 7(2):338–344, 2007.
- [27] R. Twarock. A tiling approach to virus capsid assembly explaining a structural puzzle in virology. *J. Theor. Biol.*, 226(4):477–482, 2004.
- [28] T. Keef, A. Taormina, and R. Twarock. Assembly models for papovaviridae based on tiling theory. *Phys. Biol.*, 2(3):175–188, 2005.
- [29] T. Keef, C. Micheletti, and R. Twarock. Master equation approach to the assembly of viral capsids. *J. Theor. Biol.*, 242(3):713–721, 2006.
- [30] R. Twarock. Mathematical virology: a novel approach to the structure and assembly of viruses. *Phil. Trans. R. Soc. A*, 364:3357–3373, 2006.
- [31] R. S. Schwartz, R. L. Garcea, and B. Berger. 'local rules' theory applied to polyomavirus polymorphic capsid assemblies. *Virology*, 268(2):461–470, 2000.
- [32] M. F. Hagan and D. Chandler. Dynamic pathways for viral capsid assembly. *Biophys. J.*, 91:42–54, 2006.
- [33] D. Endres, M. Miyahara, P. Moisant, and A. Zlotnick. A reaction landscape identifies the intermediates critical for self-assembly of virus capsids and other polyhedral structures. *Prot. Sci.*, 14:1518–1525, 2005.
- [34] B. Grünbaum and G. C. Shephard. *Tilings and patterns*. W. H. Freeman, New York, USA, 2nd edition, 1986.
- [35] B. Grünbaum and G. C. Shephard. Tilings with congruent tiles. *Bull. Amer. Math. Soc. N.S.*, 3:951–973, 1980.
- [36] P. W. Fowler and K. M. Rogers. Spiral codes and goldberg representations of icosahedral fullerenes and octahedral analogues. *J. Chem. Inf. Comput. Sci.*, 41(1):108–111, 2001.
- [37] J. F. Sadoc and R. Mosseri. *Geometrical frustration*. Cambridge University Press, Cambridge, UK, 1999.

- [38] T. Sugimoto and T. Ogawa. Tiling problem of convex pentagon. *Forma*, 15:75–79, 2000.
- [39] T. Sugimoto and T. Ogawa. Systematic study of convex pentagonal tilings, i: Case of convex pentagons with four equal-length edges. *Forma*, 20:1–18, 2005.
- [40] S.C. Harrison, A.J. Olson, C.E. Schutt, F.K. Winkler, and G. Bricogne. Tomato bushy stunt virus at 2.9 Å resolution. *Nature*, 276:368–373, 1978.
- [41] C. Abad-Zapatero, S.S. Abdel-Meguid, J.E. Johnson, A.G.W. Leslie, I. Rayment, M.G. Rossmann, D. Suck, and T. Tsukihara. Structure of southern bean mosaic virus at 2.8 Å resolution. *Nature*, 286:33–39, 1980.
- [42] M. G. Rossmann and J. E. Johnson. Icosahedral rna virus structure. *Annu. Rev. Biochem.*, 58:533–573, 1989.
- [43] T. Lin and J. E. Johnson. Structures of picorna-like plant viruses: implications and applications. *Adv Virus Res.*, 62:167–239, 2003.
- [44] T. Stehle, S. J. Gamblin, Y. Yan, and S. C. Harrison. The structure of simian virus 40 refined at 3.1 Å resolution. *Structure.*, 4(2):165–182, 1996.
- [45] R. C. Liddington, Y. Yan, J. Moulai, R. Sahli, T. L. Benjamin, and S. C. Harrison. Structure of simian virus 40 at 3.8-Å resolution. *Nature.*, 354(6351):278–284, 1991.
- [46] 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. Viperdb: a relational database for structural virology. *Nucl. Acids Res.*, 34:D386–D389, 2006.
- [47] J.J. Rux and R.M. Burnett. Spherical viruses. *Curr Opin Struct Biol.*, 8(2):142–149, 2008.
- [48] H.S. Subramanya, K. Gopinath, M.V. Nayudu, H.S. Savithri, and M.R. Murthy. Structure of sesbania mosaic virus at 4.7 Å resolution and partial sequence of the coat protein. *J. Mol. Biol.*, 229(1):20–25, 1993.
- [49] F. Coulibaly, C. Chevalier, I. Gutsche, J. Pous, J. Navaza, S. Bressanelli, B. Delmas, and F. A. Rey. The birnavirus crystal structure reveals structural relationships among icosahedral viruses. *Cell (Cambridge, Mass., USA)*, 120:761, 2005.
- [50] J.E. Padilla, C. Colovos, and T.O. Yeates. Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl. Acad. Sci. USA*, 98(5):2217–2221, 2001.

- [51] K.H. Choi, M.C. Morais, D.L. Anderson, and M.G. Rossmann. Determinants of bacteriophage  $\phi$ 29 head morphology. *Structure*, 14(11):1723–1727, 2006.
- [52] M. F. Moody. Geometry of phage head construction. *J. Mol. Biol.*, 293(2):401–433, 1999.
- [53] J. Lepault, I. Petitpas, I. Erk, J. Navaza, D. Bigot, M. Dona, P. Vachette, J. Cohen, and F.A. Rey. Structural polymorphism of the major capsid protein of rotavirus. *EMBO J.*, 20(7):1498–1507, 2001.
- [54] V.S. Reddy and J.E. Johnson. *Virus Structure and Assembly (Advances in Virus Research)*, chapter Structure-Derived Insights into Virus Assembly., pages 45–468. Elsevier Academic Press, San Diego, CA, USA, 2005.
- [55] D. Ferreira, R. Hernandez, M. Horton, and D.T. Brown. Morphological variants of sindbis virus produced by a mutation in the capsid protein. *Virology.*, 307(1): 54–66, 2003.
- [56] V. Sangita, P. S. Satheshkumar, H. S. Savithri, and M. R. Murthy. Structure of a mutant t=1 capsid of sesbania mosaic virus: role of water molecules in capsid architecture and integrity. *Acta Crystallogr D Biol Crystallogr.*, 61:1406–1412, 2005.
- [57] S.B. Larson, R.W. Lucas, and A.g McPherson. Crystallographic structure of the t=1 particle of brome mosaic virus. *J. Mol. Biol.*, 346(3):815–831, 2005.
- [58] C. Zubieta, G. Schoehn, J. Chroboczek, and S. Cusack. The structure of the human adenovirus 2 penton. *Mol Cell.*, 17(1):121–135, 2005.
- [59] S.J. Stray, C. Bourne, S. Punna, W.G. Lewis, M.G. Finn, and A. Zlotnick. A heteroaryldihydropyrimidine activates and can misdirect hepatitis b virus capsid assembly. *Proc. Natl. Acad. Sci. USA*, 102:8138–8143, 2005.
- [60] C. R. Bourne, M. G. Finn, and A. Zlotnick. Global structural changes in hepatitis b virus capsids induced by the assembly effector hap1. *J Virol.*, 80(22):11055–11061, 2006.
- [61] C.M. Teschke, J. King, and P. E. Prevelige Jr. Inhibition of viral capsid assembly by 1,1'-bi(4-anilinonaphthalene-5-sulfonic acid). *Biochemistry*, 32(40):10658–10665, 1993.
- [62] P.E. Prevelige Jr. Inhibiting virus-capsid assembly by altering the polymerisation pathway. *Trends Biotechnol.*, 16(2):61–65, 1998.

- [63] J. F. Conway, N. Cheng, P. D. Ross, R. W. Hendrix, R. L. Duda, and A. C. Steven. A thermally induced phase transition in a viral capsid transforms the hexamers, leaving the pentamers unchanged. *J. Struct. Biol.*, 158(2):224–232, 2007.
- [64] W. R. Wikoff, J. F. Conway, J. Tang, K. K. Lee, L. Gan, N. Cheng, R. L. Duda, R. W. Hendrix, A. C. Steven, and J. E. Johnson. Time-resolved molecular dynamics of bacteriophage hk97 capsid maturation interpreted by electron cryo-microscopy and x-ray crystallography. *J. Struct. Biol.*, 153(3):300–306, 2006.
- [65] J. Lidmar, L. Mirny, and D.R. Nelson. Virus shapes and buckling transitions in spherical shells. *Phys. Rev. E*, 68(5 Pt 1):051910, 2003.
- [66] T.T. Nguyen, R.F. Bruinsma, and W.M. Gelbart. Elasticity theory and shape transitions of viral shells. *Phys. Rev. E*, 72(5 Pt 1):051923, 2005.
- [67] F. Tama and C.L. Brooks III. Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. *J. Mol. Biol.*, 345(2):299–314, 2005.
- [68] B. Jacrot. Studies on the assembly of a spherical plant virus. ii. the mechanism of protein aggregation and virus swelling. *J. Mol. Biol.*, 95(3):433–446, 1975.
- [69] A. Helenius, E. Fries, and J. Kartenbeck. Reconstitution of semliki forest virus membrane. *J Cell Biol.*, 75(3):866–880, 1977.
- [70] B. L. Trus, F. P. Booy, W. W. Newcomb, J. C. Brown, F. L. Homa, D. R. Thomsen, and A. C. Steven. The herpes simplex virus procapsid: structure, conformational changes upon maturation, and roles of the triplex proteins vp19c and vp23 in assembly. *J. Mol. Biol.*, 263(3):447–462, 1996.
- [71] L. Li, S.M. Lok, I.M. Yu, Y. Zhang, R.J. Kuhn, J. Chen, and M.G. Rossmann. The flavivirus precursor membrane-envelope protein complex: structure and maturation. *Science.*, 319(5871):1830–1834, 2008.
- [72] V. Sangita, G. L. Lokesh, P. S. Satheshkumar, C. S. Vijay, V. Saravanan, H. S. Savithri, and M. R. Murthy. T=1 capsid structures of sesbania mosaic virus coat protein mutants: determinants of t=3 and t=1 capsid assembly. *J. Mol. Biol.*, 342(3):987–999, 2004.
- [73] F. Tama and C.L. Brooks III. The mechanism and pathway of ph induced swelling in cowpea chlorotic mottle virus. *J. Mol. Biol.*, 318(3):733–747, 2002.

- [74] J.M. Grimes, J.N. Burroughs, P. Gouet, J.M. Diprose, R. Malby, S. Zintara, P.P. Mertens, and D.I. Stuart. The atomic structure of the bluetongue virus core. *Nature.*, 395(6701):470–478, 1998.
- [75] A. Nakagawa, N. Miyazaki, J. Taka, H. Naitow, A. Ogawa, Z. Fujimoto, H. Mizuno, T. Higashi, Y. Watanabe, T. Omura, R. H. Cheng, and T. Tsukihara. The atomic structure of rice dwarf virus reveals the self-assembly mechanism of component proteins. *Structure.*, 11(10):1227–1238, 2003.
- [76] W. Jiang, Z. Li, Z. Zhang, M.L. Baker, P. E. Prevelige Jr, and W. Chiu. Coat protein fold and maturation transition of bacteriophage p22 seen at subnanometer resolutions. *Nat Struct Biol.*, 10(2):131–135, 2003.
- [77] B. Plestenjak. An algorithm for drawing planar graphs. *Softw. Pract. Exper.*, 29: 973–984, 1999.
- [78] R. Zandi, P. van der Schoot, D. Reguera, W. Kegel, and H. Reiss. Classical nucleation theory of virus capsids. *Biophys J.*, 90(6):1939–1948, 2006.
- [79] A. Zlotnick, R. Aldrich, J.M. Johnson, P. Ceres, and M.J. Young. Mechanism of capsid assembly for an icosahedral plant virus. *Virology.*, 277(2):450–456, 2000.
- [80] T. Dokland, R. McKenna, L.L. Ilag, B.R. Bowman, N.L. Incardona, B.A. Fane, and M.G. Rossmann. Structure of a viral procapsid with molecular scaffolding. *Nature.*, 389(6648):308–313, 1997.
- [81] K. Henrick, R. Newman, M. Tagari, and M. Chagoyen. Emdep: a web-based system for the deposition and validation of high-resolution electron microscopy macromolecular structural information. *J Struct Biol.*, 144(1-2):228–237, 2003.
- [82] Thomas C Mettenleiter, Barbara G Klupp, and Harald Granzow. Herpesvirus assembly: an update. *Virus Res*, 143(2):222–234, Aug 2009. doi: 10.1016/j.virusres.2009.03.018. URL <http://dx.doi.org/10.1016/j.virusres.2009.03.018>.
- [83] H.T Jääliñoja, E. Roine, P. Laurinmäki, H.M. Kivelä, D.H. Bamford, and S.J. Butcher. Structure and host-cell interaction of sh1, a membrane-containing, halophilic euryarchaeal virus. *Proc. Natl. Acad. Sci. USA*, 105(23):8008–8013, 2008.
- [84] A.K. Overby, R.F. Pettersson, K. Grünewald, and J.T. Huiskonen. Insights into bunyavirus architecture from electron cryotomography of uukuniemi virus. *Proc. Natl. Acad. Sci. USA*, 105(7):2375–2379, 2008.

- [85] S. Libersou, X. Siebert, M. Ouldali, L.F. Estrozi, J. Navaza, A. Charpilienne, P. Garnier, D. Poncet, and J. Lepault. Geometric mismatches within the concentric layers of rotavirus particles: a potential regulatory switch of viral particle transcription activity. *J Virol.*, 82(6):2844–2852, 2008.
- [86] A. Fokine, V.A. Kostyuchenko, A.V. Efimov, L.P. Kurochkina, N.N. Sykilinda, J. Robben, G. Volckaert, A. Hoenger, P.R. Chipman, A.J. Battisti, M.G. Rossmann, and V.V. Mesyanzhinov. A three-dimensional cryo-electron microscopy structure of the bacteriophage phikz head. *J Mol Biol.*, 352(1):117–124, 2005.
- [87] Juha T Huiskonen, Anna K Overby, Friedemann Weber, and Kay Gr̃unewald. Electron cryo-microscopy and single-particle averaging of rift valley fever virus: evidence for gn-gc glycoprotein heterodimers. *J Virol*, 83(8):3762–3769, Apr 2009. doi: 10.1128/JVI.02483-08. URL <http://dx.doi.org/10.1128/JVI.02483-08>.
- [88] X. Yan, P.R. Chipman, T. Castberg, G. Bratbak, and T.S. Baker. The marine algal virus ppv01 has an icosahedral capsid with  $t=219$  quasisymmetry. *J Virol.*, 79(14):9236–9243, 2005.
- [89] X. Yan, N.H. Olson, J.L. Van Etten, M. Bergoin, M.G. Rossmann, and T.S. Baker. Structure and assembly of large lipid-containing dsdna viruses. *Nat Struct Biol.*, 7(2):101–103, 2000.
- [90] C.C. Lin and L.A. Segel. *Mathematics applied to deterministic problems in the natural sciences*, chapter Field Equations of Continuum Mechanics., pages 490–491. SIAM, Philadelphia, USA, 1988.
- [91] J.F. Conway, W.R. Wikoff, N. Cheng, R.L. Duda, R.W. Hendrix, J.E. Johnson, and A.C. Steven. Virus maturation involving large subunit rotations and local refolding. *Science.*, 292(5517):744–748, 2001.
- [92] B.W. Neuman, B.D. Adair, J.W. Burns, R.A. Milligan, M.J. Buchmeier, and M. Yeager. Complementarity in the supramolecular design of arenaviruses and retroviruses revealed by electron cryomicroscopy and image analysis. *J Virol.*, 79(6):3822–3830, 2005.
- [93] Tam Luong Nguyen, Guy Schoehn, Winfried Weissenhorn, Ann R Hermone, James C Burnett, Rekha G Panchal, Connor McGrath, Dan W Zaharevitz, M. Javad Aman, Rick Gussio, and Sina Bavari. An all-atom model of the pore-like

---

structure of hexameric vp40 from ebola: structural insights into the monomer-hexamer transition. *J Struct Biol*, 151(1):30–40, Jul 2005. doi: 10.1016/j.jsb.2005.02.013. URL <http://dx.doi.org/10.1016/j.jsb.2005.02.013>.

# Appendix A

## Appendix: Capsid Geometry

This is the supporting material to Chapter 4.

### A.1 Canonical intra-pentameric dihedral angles interact at $\sim 138.19^\circ$

Hexamers and pentamers within a canonical capsid (those capsids representable as “monohedral tilings” that display few holes, few overlaps and structural invariability; please see Section 3.1.1.3 for a more rigorous description) may be treated as a six- and five-coordinated set of plates respectively (Figure A.1A, B).

**Statement:** Any canonical subunit that possesses the ability to form both pentamer and hexamer must possess intra-pentamer dihedral angles of  $\sim 138.19^\circ$  (i.e., in Figure A.1D,  $\phi \sim 138.19^\circ$ ).

From the canonical capsid definitions (Mannige and Brooks, 2008), we get that (1) subunits can form both flat hexamers (Figure A.1B) and “curved” pentamers (Fig-

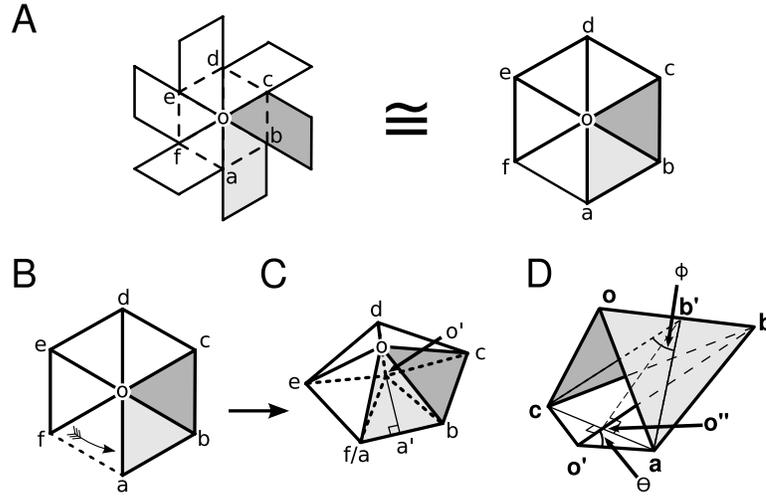


FIGURE A.1: **Showing that  $\phi \sim 138.19^\circ$ .** (A) indicates that n-valent clusters formed from trapezoids (shown in the diagram for hexamers) may be reduced/simplified to clusters of equilateral triangles for the purpose of analyzing dihedral angle properties. Hexamers (B) and pentamers (C) in canonical capsids are formed from the same subunit interface that interact at varying dihedral angles. A pair of adjacent subunits is shaded in in the pentamer (C) and isolated environment (D), which will be used to obtain a relationship for the dihedral angle ( $\phi$ ).

ure A.1C) from the same interface, and (2) all angles within a pentamer are identical (which is a reasonable assumption given that i. these pentameric angles are formed from identical interfaces and not quasi-equivalent ones, and ii. a five-fold rotational symmetry element (axis) falls perpendicular to the center of pentamer in the crystal structure).

**Proof:** Specifically, from the right triangle  $b'o''a$  in Figure A.1D, we can obtain a relationship for the intra-pentamer dihedral angle ( $\phi$ ) and edge lengths:

$$\sin(\phi/2) = \frac{|o''a|}{|ab'|} \quad (\text{A.1})$$

We now assume that the edge of the equilateral triangular faces is 1 with no loss of generality. Given that right triangle  $ab'o$  is a 30-60-90 triangle and that  $|oa| = 1$ , we

get

$$|ab'| = \sqrt{3}/2 \quad (\text{A.2})$$

Substituting Equation A.2 in Equation A.1 and rearranging, we get,

$$\sin(\phi/2) = \frac{2}{\sqrt{3}}|o''a| \quad (\text{A.3})$$

Theta ( $\theta$ ) is the angle of the radiating edges once projected to a plane that contains a,b,c,d,e (there must be such a plane since all dihedral angles are set to the same value). From Figure A.1C, we get

$$|o''a| = |o'a| \sin(\theta) \quad (\text{A.4})$$

Substituting Equation A.4 into Equation A.3, we get

$$\sin(\phi/2) = \frac{2}{\sqrt{3}}|o'a| \sin(\theta) \quad (\text{A.5})$$

Also, from the 30-60-90 triangle  $o'aa'$  in Figure A.1B, we get

$$|o'a| = \frac{|a'a|}{\sin(\theta/2)} = \frac{1}{2 \sin(\theta/2)} \quad (\text{A.6})$$

Substituting Equation A.6 in Equation A.5, we get

$$\sin(\phi/2) = \frac{\sin(\theta)}{\sqrt{3} \sin(\theta/2)} \quad (\text{A.7})$$

or

$$\phi = 2a \sin \left( \frac{\sin(\theta)}{\sqrt{3} \sin(\theta/2)} \right) \quad (\text{A.8})$$

If all the dihedral angles within the pentamer are alike, then  $\theta = 2\pi/5$  (this generalizes to  $\theta = 2\pi/i$  if the pentamer is actually an  $i$ -mer), and

$$\phi = 2a \sin \left( \frac{\sin(2\pi/5)}{\sqrt{3} \sin(\pi/5)} \right) \sim 138.19^\circ \quad (\text{A.9})$$

This will be true for any set of canonical capsid subunits that assemble into pentamers, and is also seen in true icosahedra (twenty faced deltahedra) that describe  $T = 1$  capsids, which, we claim, allows for  $T > 1$  to  $T = 1$  transformations (see main text).

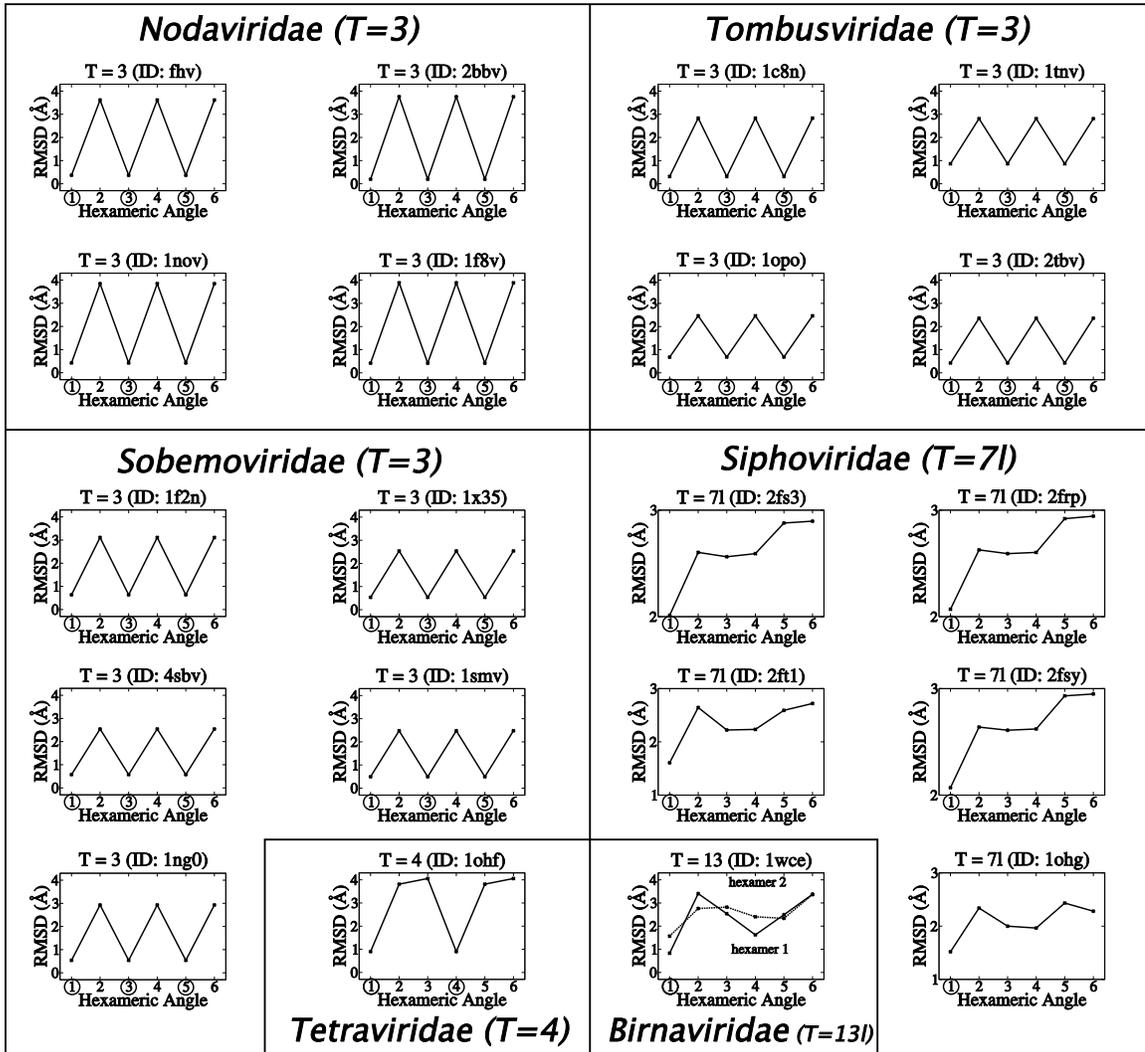


FIGURE A.2: **Hexamer angle profiles.** Angle profiles of hexamers in unique environments when compared to a pentameric endo angle of the same capsid (low RMSD values indicate more pentamer-like angles) shown for individual capsids (indicated by their PDBID or ID). This graph is an expanded version of Figure 4.2A.

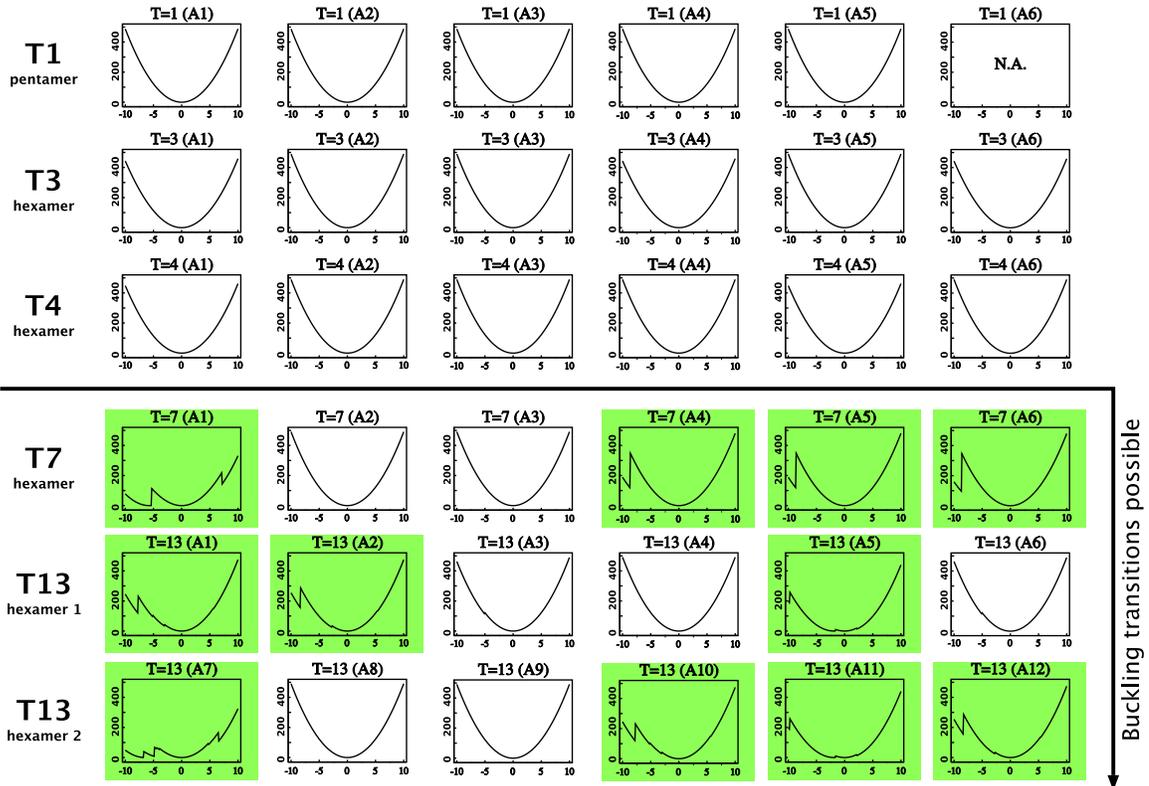


FIGURE A.3: **Capsid model buckling profiles.** Energy (y-axis) vs. angle constraint disequilibrium ( $r_0 - r'$ ; x-axis) profiles for individual angles (labeled A1 thru A6 for a unique hexamer and an additional A7 thru A12 for the second hexamer) within hexamers for capsids with  $T = 3, 4, 7, 13$ .  $T = 1$  pentameric angle profiles are included to give a sense of rigid angle profiles. Only  $T > 4$  canonical capsids hexamer angles (whose profiles are highlighted green) appear to sample multiple conformations upon application of small forces indicating that buckling transitions are possible for only  $T \geq 4$  canonical capsids. This figure is an expansion of Figure 4.4B in the main text.

## Appendix B

# Appendix: Capsid Periodic Table

**This text is the supplementary material to Chapter 5 and performs the following functions:**

- ✓ Defines hexamer complexity (Section B.2) and shape (Section B.3).
- ✓ Provides additional data on capsid abundances (Section B.5, Figure B.2) further indicating the utility of hexamer complexity (Section B.2) and the periodic table (Figure 5.5C) in explaining evolutionary pressures (Section B.4).
- ✓ Critically evaluates the validity of our results (Section B.6).
- ✓ Provides a formalism for calculating hexamer complexity (Sections B.7-B.10).

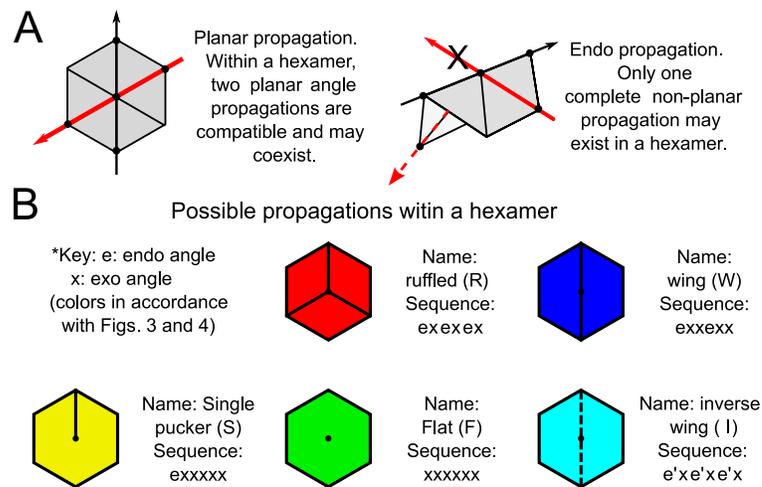
### B.1 Canonical vs. noncanonical capsids

All our specific predictions are directed towards canonical capsids where subunits (within any given capsid) are tilable and nearly-invariant in shape [1]. This is because the consequence of introducing/imposing curvature into the shell is conveniently imposed as endo angle propagations [2], which then allows for hexamer shapes to be

precisely characterized (Section B.2). However, that our predictions apply to all structurally characterized spherical capsids indicate parallel constraints applied to non-canonical capsid hexamers. It will be interesting to see the differences and similarities between the constraints acting on canonical and noncanonical capsids.

## B.2 Defining hexamer complexity $C^h$

Hexamer complexity  $C^h$  is the minimal number of distinct hexamer shapes that a canonical capsid [1] of specific size (defined by  $h, k$  or  $T$ ) contains. The possible hexamer shapes that a canonical capsid may possess are shown in Figure B.1B (derived by inspecting Fig. 2 and assuming the working of endo angle propagation and termination rules in Figure 5.2).



**FIGURE B.1: Hexamer shapes available to capsids.** (A) Although planar endo angle constraints are able to freely propagate within hexamers (left), only one *complete* non-planar (or “endo”) angle constraint/propagation may be present within a hexamer (collinear propagations not included). If two non-linear/non-parallel propagations meet, one must terminate at that meeting point, which means that multiple non-linear endo angles may exist within a single hexamer only if *terminated* at its center. (B) Possible arrangements of terminal endo angles (reflecting possible hexamer shapes) are listed (endo angles are represented as lines in the hexamer diagrams and as e in the hexamer angle sequence; e' represents an inverse endo angle). The hexamers are colored in accordance with the Figures 5.3 and 5.2.

### B.3 Counting hexamer shapes

In Section 4.2.1, we showed that different arrangements of **endo** dihedral angles (designated “e”) among non-endo, or **exo** angles (designated “x”) in a hexamer define distinct hexamer shapes [2]. This assumption has been shown to be true for those natural canonical capsids that have afforded investigation [2]; specifically, we showed that the smallest capsids from each class ( $T = 3, 4, 7$ ) possess distinct hexamer shapes, named in accordance with the hexamer coloring in Figs. 2 and 3: red (*exexex*; “ruffled”), blue (*exxxx*; “wing shaped”), and yellow (*xxxxxx*, “single-pucker”) hexamer shapes respectively [2]. These capsids possess the lowest  $C^h$  of one.

Larger capsids increase in  $C^h$  due to the requirement of additional hexamer shapes colored in Fig. 2 as green (*xxxxxx*; “flat”<sup>1</sup>) and cyan (*e'xxe'xx*, shaped as an “inverse wing” possessing *inverse endo angles* e' whose acute angles face outward).

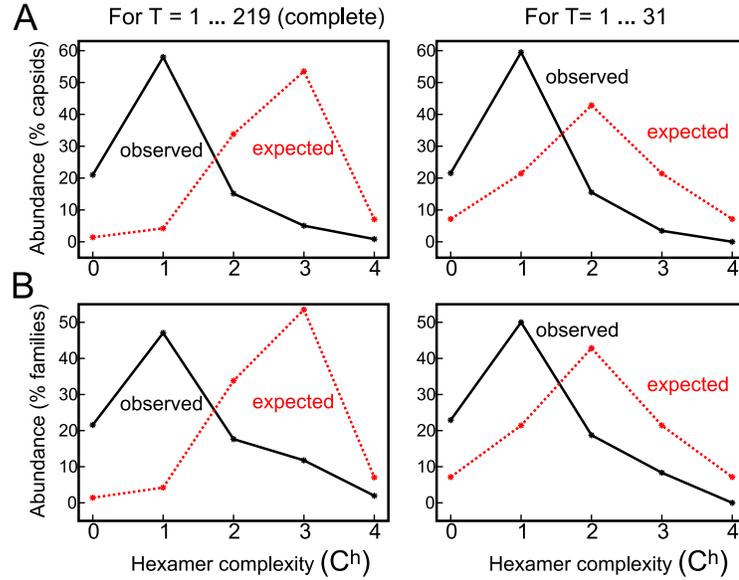
### B.4 Capsids with low $C^h$ are preferred

From Figure B.2, we can surmise that, for the range of  $T$  numbers observed ( $T = 1...219$  and for a more conservative/truncated range,  $T = 1...31$ ), capsids with lower  $C^h$  appear to be preferred as evidenced by a shift to lower  $C^h$  distributions in observed versus expected capsid distributions. Table B.1 lists the first twelve capsid sizes ( $T$ ) by class; those sizes displaying  $C^h > 2$  are indicated by boldface.

A major difference between the red and black graphs in Figure B.2 comes in the behavior in abundances of expected  $C^h = 3$  capsids, that mostly belong to the  $h > k > 1$

---

<sup>1</sup>In the  $h > k = 1$  capsids, the green hexamer is not perfectly flat, but will tend towards possessing identical dihedral angles, which, for a hexamer, optimally would result in generally flat hexamers.



**FIGURE B.2: Capsids tend to prefer lower  $C^h$  than expected.** Plotted in each graph is  $C^h$  versus observed (solid, black lines) and expected abundances (dotted, red lines) obtained from 119 capsids (A) and 52 families (B) shown for the complete available capsid size range ( $T = 1 \dots 219$ ; left) and a truncated range (right). The expected dataset assumed that no capsid size ( $T$ ) is preferred (i.e., it assumes a uniform capsid size distribution among viruses). Families that display two capsid sizes were split to maintain one  $C^h$  per family entry. This figure is an expansion of Figure 5.4A.

regime. Specifically, as we increase from the  $(n-1)^{th}$  period to  $n^{th}$  period in the periodic table, class 1 (where  $C^h$  mostly equals 2) and class 3 entries (where  $C^h$  mostly equals 4) increase by 1, while the class 2 entries (where  $C^h$  mostly equals 3) increase in a more-or-less arithmetic progression by  $(n-1)$  (evident in Fig 3C in the triangular shape of the class 2 group versus linear shapes of class 1 and class 3 groups respectively).

**TABLE B.1: The distribution of capsid sizes into the three morphological classes described by the relationship between the capsid's  $h$  and  $k$ .** The percentage abundance ( $A(\%)$ ) of capsids in the three classes were obtained from a collection of 118 non-redundant capsids belonging to 39 diverse capsid families.

Class	$h-k$	$A(\%)$	Triangulation ( $T$ ) number series						
1	$h > k = 0$	33.9	1	4	9	16	25	...	
2	$h > k > 0$	22.8			7	13	19	21	...
3	$h = k$	43.2	3		12			27	...

## B.5 Observed capsid abundance $\propto 1/C^h$

Finally, excepting  $C^h = 0$  capsids (i.e., capsids that contain no hexamers, or  $T = 1$  capsids), there is an inverse relationship between  $C^h$  and observed capsid abundance (black lines in Figure B.2). The low observed abundance for  $C^h = 0$  capsids is expected, given that most virus families with true  $C^h = 0$  appear to be too small to accommodate enough genomic material to infect as a primary source (therefore, most true  $T = 1$  capsids belong to “satellite viruses” that are only able to infect hosts pre-infected by a primary infector, presumably since those virus capsids provide insufficient volume to contain an independent infectious genome). Here, the additional/stronger evolutionary impediment appears to be a lower bounded genome size preference (i.e., a non-geometric preference imposing a constraint of  $C^h > 0$  may be overlaid with the inverse  $C^h$  rule to obtain the observed or black graphs in Figure B.2).

## B.6 Is there a data-collection bias?

Here, we address the question: are our findings a result of a basic inability to sample structures of large  $C^h$ , or does the data truly reflect our predictions?

Figure B.2 (reflecting the rest of our data) was produced from a compilation of capsids obtained from (1) X-ray crystallography, whose prowess lies in obtaining high resolution capsid 3D structures of “small” sizes (e.g.,  $T = 1...25$ ), and (2) electron microscopy, where large capsids do not disallow the elucidation of capsid size or  $T$  number (which can be obtained from simple electron micrographs, if not by 3D capsid reconstructions). Consequently, we argue that if observable to a structural virologist, any

new capsid of any size would not be far from finding a public domain home (thereby finding its way in our graphs). Thus we argue that our observed data does not reflect discrepancies in data collection as much as it lends credence to our geometric predictions.

Furthermore, if capsid collection were to be size constrained, it would still not matter so much, since our existence rules are not size dependent as much as  $h, k$  dependent (e.g., although smaller than  $T = 25$ , the  $T = 19$  capsid is expected to be higher in hexamer complexity and therefore lower in abundance, which is the case).

## B.7 Basic definitions

The Kronecker delta function ( $\delta_x$ ) is quite integral to our future formalisms, and is therefore introduced here as a special topic. Specifically,  $\delta_x$  (or  $\delta_{x,0}$ ) is an algorithm, that outputs 1 if  $x=0$  and 0 otherwise, i.e.,

$$\delta_x = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases} \quad (\text{B.1})$$

We can represent this algorithm by the limits

$$\delta_x = \lim_{\alpha \rightarrow \infty} \frac{1}{e^{\alpha x^2}} \quad (\text{B.2})$$

or

$$\delta_x = \lim_{\alpha \rightarrow \infty} \frac{2e^{\alpha x}}{1 + e^{2\alpha x}} \quad (\text{B.3})$$

which may be used later on.

We also utilize a convenient equation that produces a binary output after comparing two non-negative integers  $a$  and  $b$ :

$$\Delta_{a>b} = \prod_{i=0}^b (1 - \delta_{(a-i)}) = \begin{cases} 1 & \text{if } a > b \\ 0 & \text{otherwise} \end{cases} \quad (\text{B.4})$$

Some basic definitions:

$$\delta_a (1 - \delta_a) = \begin{cases} 1 \times (1 - 1) & \text{if } a = 0 \\ 0 \times (1 - 0) & \text{otherwise} \end{cases} \quad (\text{B.5})$$

i.e., for all cases,

$$\delta_a (1 - \delta_a) = 0 \quad (\text{B.6})$$

Also, it follows that

$$a\delta_a = 0 \quad (\text{B.7})$$

and

$$\begin{aligned} \delta_a \Delta_{a>b} &= \delta_a (1 - \delta_a) \times [(1 - \delta_{a-1}) \dots (1 - \delta_{a-b})] \\ &= 0 \times [(1 - \delta_{a-1}) \dots (1 - \delta_{a-b})] \\ &= 0 \end{aligned} \quad (\text{B.8})$$

## B.8 Obtaining endo propagation length $\phi_{h,k}$

*Definition of endo angle propagation length  $\phi_{h,k}$ .* It is the distance (in capsomers, including the originating pentameric angle) that the endo angle is allowed to propagate

from a pentamer into the hexamers before being intercepted (or terminated). Please refer to Figure 5.2 for a review of the endo angle propagation and termination rules.

From Figure 5.2C, we can obtain the endo angle propagation length for a capsid of size  $h, k$ :

$$\phi_{h,k} = \begin{cases} h & \text{for class 1, i.e., if } k = 0 \\ k & \text{otherwise, i.e., if } k \neq 0 \end{cases} \quad (\text{B.9})$$

which can be described as

$$\phi_{h,k} = (h\delta_k + k) = \begin{cases} h & \text{if } k = 0 \\ k & \text{otherwise} \end{cases} \quad (\text{B.10})$$

## B.9 Obtaining $C^h$ from $h, k$ and $\phi_{h,k}$

Here, we obtain a mathematical/algorithmic expression for  $C^h$ . We can treat the hexamer complexity  $C^h$  as a sum of its components  $C_X^h$ , where  $X$  may be one of the five distinct hexamer shapes (i.e.,  $X \in (W, R, S, F, I)$ ), and  $C_X^h = 1$  only if the hexamer shape “ $X$ ” exists within the capsid. We now attempt to obtain the  $C^h$  components for each hexamer shape.

**Wing shaped (W).** The presence of two linear adjacent endo angles within a hexamer automatically indicate that wing shaped hexamers must exist within the capsid, since the only hexamer that can accommodate two linear angles is the winged shape of profile **exxexx** [2] (we define a *linearly adjacent angle set* as a set of two angles within the hexamer of position  $i$  and  $i + 3$ , where  $i = i + 6$ , indicating the cyclic nature of the angles). Therefore, we will expect wing shaped hexamers when  $\phi_{h,k} > 1$ . So, the

hexamer complexity contribution by the presence of a wing shaped hexamer will be

$$C_W^h = \Delta_{((h\delta_k+k)>1)} \quad (\text{B.11})$$

***Single pucker shaped (S).*** We can define the closest distance (in capsomer units) between two adjacent pentamers ( $P_{h,k}$ ) as

$$P_{h,k} = (h + k) \quad (\text{B.12})$$

Which is an interesting value, since it is also the maximum number of capsomers that the endo angle can propagate through, i.e.,

$$P_{h,k} \geq \phi_{h,k} \quad (\text{B.13})$$

We can also show that if  $\phi_{h,k} \geq P_{h,k}/2$ , then the endo angles will form a complete/unbroken cage around the capsid (which is seen in classes 1 and 3). However, if we do not have “complete propagation”, then we are guaranteed the existence of a single pucker hexamer, i.e.,

$$\begin{aligned} C_S^h &= \Delta_{(P_{h,k}>2\phi_{h,k})} \\ &= \Delta_{((h+k)>(2h\delta_k+2k))} \\ &= \Delta_{(h>(2h\delta_k+k))} \end{aligned} \quad (\text{B.14})$$

***Ruffle shaped.*** We also know that if  $h = k$  (class 3) then  $P_{h,k} = 2\phi_{h,k}$  (because if  $h = k$  then  $2\phi_{h,k} = 2h = h + k = P_{h,k}/2$ ) and three adjacent endo angles will terminate

TABLE B.2: Hexamer complexity,  $C^h$  from Equation B.18.

$T$	$h$	$k$	$C_W^h$	$+ C_S^h$	$+ C_R^h$	$+ C_I^h$	$+ C_F^h$	$= C^h$
1	1	0	0	0	0	0	0	0
3	1	1	0	0	1	0	0	1
4	2	0	1	0	0	0	0	1
7	2	1	0	1	0	0	0	1
9	3	0	1	0	0	0	1	2
12	2	2	1	0	1	1	0	3
13	3	1	0	1	0	0	1	2
16	4	0	1	0	0	0	1	2
19	3	2	1	1	0	0	1	3
21	4	1	0	1	0	0	1	2
25	5	0	1	0	0	0	1	2
27	3	3	1	0	1	1	1	4
28	4	2	1	1	0	0	1	3
31	5	1	0	1	0	0	1	2
36	6	0	1	0	0	0	1	2

at the central hexamer causing the presence a hexamer of *exexex* profile and of ruffled shape, so

$$C_R^h = \delta_{(h-k)} \quad (\text{B.15})$$

***Inverse-wing shaped.*** We know that the ruffled *exexex* profile is rigid [2], so even the exo ( $x$ ) to must remain constrained. Since this dihedral's acute angle faces the outside portion of the capsid, we call this special angle the inverse endo ( $e'$ ) angle. Since inverse endo angles are constrained, they must propagate between any two ruffled hexamers, resulting in the formation of a special inverse-wing shape in large enough capsids ( $h, k > 1$ ) containing ruffled hexamers ( $h = k$ ), i.e., we have

$$C_I^h = C_R^h \Delta_{(h>1)} = \delta_{(h-k)} \Delta_{(h>1)} = \delta_{(h-k)} \Delta_{(k>1)} \quad (\text{B.16})$$

**Flat shaped.** Finally, we know that a capsid of large enough size ( $h > 2$ ) irrespective of class, must possess hexamers that are generally unaffected by endo angle constraints which are therefore generally flat, so

$$C_F^h = \Delta_{(h>2)} \quad (\text{B.17})$$

Combining the above  $C^h$  components, our resulting relationship for hexamer complexity will be

$$\begin{aligned} C^h &= C_W^h + C_S^h + C_R^h + C_I^h + C_F^h \\ &= \Delta_{((h\delta_k+k)>1)} + \Delta_{(h>(2h\delta_k+k))} + \delta_{(h-k)} \\ &\quad + \delta_{(h-k)}\Delta_{(k>1)} + \delta_{(h-k)}\Delta_{(k>1)} + \Delta_{(h>2)} \end{aligned} \quad (\text{B.18})$$

## B.10 The number of hexamers $N_X$

We list the number of hexamers  $N_X$  per hexamer type  $X$ :

$$N_W = \left( \frac{60(h\delta_k + k)}{1 + \delta_k} \right) C_W^h \quad (\text{B.19})$$

$$N_S = 60C_S^h \quad (\text{B.20})$$

$$N_R = 20C_R^h \quad (\text{B.21})$$

$$N_I = (h - 1)C_I^h \quad (\text{B.22})$$

$$N_F = \left( 10(T - 1) - \sum_{X \in [W, S, R, I]} N_X \right) C_F^h \quad (\text{B.23})$$

## B.11 The viruses used in calculating capsid abundance in

### Chapter 5

The capsids used to calculate capsid abundances (for each  $C^h$ ), e.g., in Figures 5.4A and B.2, were obtained primarily from the virus structure databases VIPER EMDB [46] and EMDB [81] for EM structures and VIPERdb [46] for X-ray structures; due to space constraints, please refer to the supplementary material section of (Mannige and Brooks, 2010, PLoS ONE, in press) for a more detailed description of the virus capsid structures in table form. Here, we abbreviate that information by providing the virus families followed by the number of distinct viruses (in brackets) listed in order of their triangulation ( $T$ ) number and  $(h,k)$  indices ( $pT$  stands for pseudo-triangulation, which means that all subunits are not chemically identical):

**T=1 (1,0):** Adenoviridae (1), Birnaviridae (1), Bromoviridae (1), Microviridae (5), Papillomaviridae (1), Parvoviridae (8), Reoviridae (1), Satellites (3), Sobemoviridae (1), Totiviridae (2); **T=1,pT=2 (1,0):** Partitiviridae (1), Cystoviridae (2); **T=3 (1,1):** Bromoviridae (4), Caliciviridae (2), Flaviviridae (4), Leviviridae (5), Luteoviridae (1), Nodaviridae (3), Podoviridae (1), Sobemoviridae (5), Thermococcaceae (1), Tombusviridae (3), Tymoviridae (3); **pT=3 (1,1):** Comoviridae (5), Picornaviridae (10), Unknown4 [Kelp fly virus] (1); **T=4 (2,0):** Hepadnavirus (1), Podoviridae (1), Tetraviridae (2), Togaviridae (4); **T=7 (2,1):** Caulimoviridae (1), Papillomaviridae (2), Polyomaviridae (2), Podoviridae (4), Siphoviridae (2), Unknown3 [epsilon15] (1); **T=12 (2,2):** Bunyaviridae (1); **T=13 (3,1):** Birnaviridae (1), Cystoviridae (2), Reoviridae (6); **T=16 (4,0):** Herpesviridae (1); **T=19 (3,2):** Reoviridae (2); **pT=21 (4,1):** Corticoviridae (1); **pT=25**

---

**(5,1):** Adenoviridae (3), Tectiviridae (2); **T=27 (3,3):** Caudovirales (1); **T=28 (4,2):** Unknown1 [SH1] (1); **T=31 (5,1):** Unknown2 [STIV] (1); **T=147 (7,7):** Iridoviridae (1); **T=169 (8,7):** Phycodnaviridae (1); **T=210 (10,7):** Phycodnaviridae (1).

SO LONG, AND THANKS FOR ALL THE FISH!