

# Capsids and Stuff

Charles Zahara

June 6, 2015

# Contents

<b>List of Figures</b>	<b>3</b>
<b>1 Introduction</b>	<b>4</b>
<b>2 Background</b>	<b>5</b>
2.1 The History of Icosahedral Symmetry . . . . .	7
2.1.1 Initial Theories of Crick and Watson . . . . .	7
2.1.2 The Theory of Quasi-Equivalence . . . . .	10
2.1.3 Recent Work . . . . .	12
<b>3 Mathematical Problems and Solutions</b>	<b>12</b>
3.1 Triangulation Numbers . . . . .	12
3.1.1 Solution 1: Geometry . . . . .	13
3.1.2 Protein Shape . . . . .	18
3.1.3 Protein Shape . . . . .	18
<b>4 Conclusion</b>	<b>20</b>
<b>5 References</b>	<b>21</b>

## List of Figures

1	Viral life cycle . . . . .	6
2	Viral life cycle . . . . .	7
3	Symmetry examples. Use figures from Mannige . . . . .	9
4	Table of cubic point groups . . . . .	10
5	Equilateral Triangle on h, k-axes . . . . .	14
6	Area of Equilateral Triangle Constructions . . . . .	15
7	. . . . .	16
8	. . . . .	16
9	. . . . .	17

# 1 Introduction

Mathematics has always been a beautiful field full of excitement and discovery. From purely theoretical considerations to strictly applied problems the mathematician is never without mental stimulus. Its inspirations are everywhere and one needs only to be willing to ask why and how to be able to find a fantastic problem to consider. Although the sciences sometimes lack a willingness to work together, Biology is a great realm within which to discover new questions. To find challenging mathematical problems within Biology, one does not need to be an applied mathematician with decades of experience in modeling. Nor does one need to have an extensive background in Biology and science. All that is required is to be willing to examine systems and be willing to ask questions.

The aim of this paper is to take the reader through the discovery of mathematical puzzles inspired by the physical world. Specifically these mathematical questions will come from the theory of spherical virus capsids. Two problems left unanswered by many of the works on the spherical virus capsids will be laid out and then multiple techniques for solving these problems will be detailed. Although this paper is centered around the mathematics of why icosahedral shells work, it aims to be readable by anyone without external materials, extensive background reading, or a 6' by 20' wall of white board to work out gaps in equations. This paper will ultimately demonstrate that unique problems in mathematics are everywhere and can often be solved with very basic tools if one is willing to apply some hard work and problem solving.

A secondary objective of this paper is to be an entertaining introduction to the theory of icosahedral symmetry of virus capsids. It is intended to be a type of tour guide through the development and arguments of this theory. As such, the paper begins with a review of both past and present work on virus capsids. First the pivotal works on the theory of symmetry in spherical virus capsids will be summarized. Then a few selections of the

current work both biologists and mathematicians are performing in the field of virus capsids will be examined. Hopefully the reader will come away understanding not only the form and function of icosahedral capsids, but also the amazing, challenging, and ultimately fun mathematical problems that arise within Biology.

## 2 Background

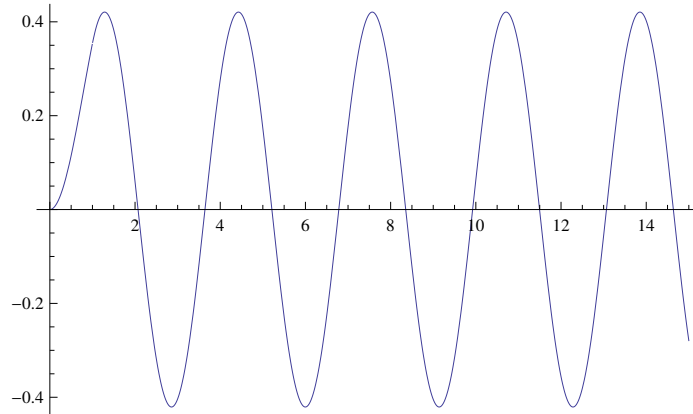
Viruses have been shaping human history for millennia. From smallpox and polio to the yearly strains of flu and common cold, every generation has battled against these tiny creatures. They have wiped out civilizations and continue to kill millions every year. As such the study of their form and function remains important.

**Definition** (Virus). *Viruses are infectious agents composed of nucleic acid surrounded by a protein shell, called a capsid, and in some cases a membranous envelope.* [1]

While the above definition may make viruses seem like simple little organisms, they have a multitude of unique variations. Their nucleic acid genetic material may come in the form of DNA or RNA, depending on the virus, and be either single or double stranded. They can be rod shaped, roughly spherical, or in some cases rather complex in shape. Additionally, viruses are very effective organisms. They cannot replicate or reproduce themselves, but instead inject their genetic material into a host cell. The host cell is then taken over and forced to create copies of the virus genetic material and capsid proteins. Eventually, once it is filled to capacity, the host cell bursts open and the newly created capsids are expelled and proceed to a new host to repeat the process (See Figure 1). [1]

The purpose of these capsid shells are to protect the virus genetic material as they move from host cell to host cell. For instance, it has been shown for some virus particles that the capsid shell protects the nucleic acid from ribonuclease and temperature changes [2, p 5]. This makes the capsid crucial in the infection process. It also means that understanding the

Figure 1: Viral life cycle



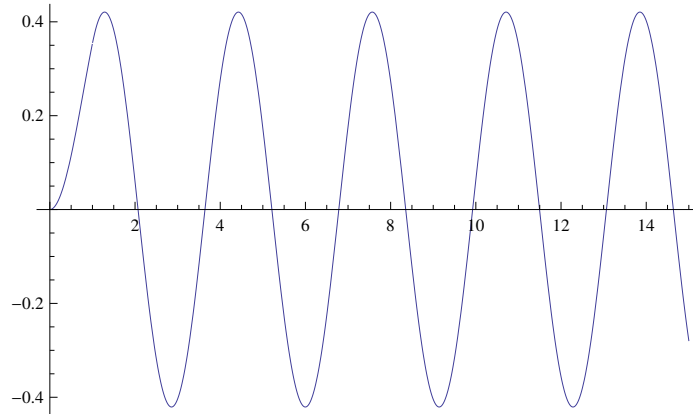
structure and mechanics of the capsids may allow scientists to develop ways to disrupt their construction, thus eliminating the virus's ability to travel and infect new host cells.

As stated previously, viruses come in a variety of shapes and sizes, defined by their capsid shells. Some are long tubes, similar to a straw, classified as helical viruses, others are roughly spherical, classified as icosahedral viruses, and others do not fall into either of the two previously mentioned categories and are classified as complex (See Figure 2). Spherical viruses are the focus of this paper. In the late 1950's and early 1960's it was suggested that spherical viruses capsids must conform to icosahedral symmetry [3], [2]. Although there are still many exceptions which continue to be examined, through decades of study and advances in microscope technology the theory that spherical capsids must have icosahedral symmetry has been confirmed both visually and computationally for most cases. What is often left unexplained in papers discussing spherical viruses is the mathematical reasoning behind the theory of their shape.

Before the history and theory are reviewed a few key definitions and their relationship are important.

**Definition** (Capsid). *A viral capsid is a protein shell made up multiple copies of one, but in some cases a few, small proteins, which encloses and protects the viral genome.*

Figure 2: Viral life cycle



**Definition** (Capsomere). *A capsomere is a repeated symmetric building unit for the viral capsid. They are made up of multiple proteins.*

**Definition** (Protomer). *A protomer is a single protein unit of a capsomere.*

The important note is that the protomers are the smallest unit. They are single asymmetric proteins. Capsomeres are built up of multiple protomers in a symmetric fashion. There can be multiple differently shaped capsomeres made from the same protomer subunit. The Capsid is the final shell, made up of the capsomere building blocks.

## 2.1 The History of Icosahedral Symmetry

### 2.1.1 Initial Theories of Crick and Watson

In 1956 F. H. Crick and J. D. Watson wrote a paper entitled "Structure of Small Viruses." In this paper they hypothesized that "a small virus contains identical sub-units, packed together in a regular manner" [3, p 473]. While this was not an entirely new idea, Crick and Watson were the first to describe the specifics of the idea and to proclaim it as a feature of all small viruses. This hypothesis was based upon the fact that all plant viruses having been studied at the time were extremely "regular," meaning that for a particular virus each

particle was similar, if not identical, in appearance. This belief was strengthened by the fact that viruses can form crystals, a feature only possible if each individual viral molecule is identical.

At the time of their writing, viruses had only been observed to follow the rod shape of the tobacco-mosaic virus (TMV), as studied by Tubingen and Berkley, or the spherical shape of the turnip yellow mosaic virus (TYMV), as studied by Markham. Indeed, more research had been done on the Tobacco Mosaic Virus than any spherical virus at the time and so many conjectures were made from these observations. For instance, the tobacco mosaic virus consists of 94% protein and 6% RNA [3, p 473]. This suggests that proteins must be repeated as there is not enough RNA to code unique proteins in that ratio. Very early X-Ray work showed that the each TMV particle was made up of sub-units of some sort. More importantly, Crick and Watson were able to observe that the TMV particles had a screw like axis, which acted as a symmetry axis, thus placing each sub-unit in an identical environment as the rest.

While Crick and Watson never used the words capsomere and protomer in their paper, it is beneficial to differentiate them here. In the case of the TMV, the sub-units that Crick and Watson were speaking of were actually the individual protein units, the protomeres. This is evidenced by their description of the sub-unit as the "small globular proteins" [3, p 473]. In the case of helical viruses there are actually no capsomeres, the repeated sub-unit packed around a helical axis are simply the individual proteins themselves [3, p 474]. It is helpful to understand what Crick and Watson were observing on the TMV particles in order to understand the conjectures they made about spherical virus.

Crick and Watson extended their observations of the TMV particles to the less studied spherical viruses. It had been shown that TYMV and bushy stunt virus would crystalize into a unit cell that had the shape of a cube. This forces the individual virus particles to have cubic symmetry. Indeed, Crick and Watson were in communication with Caspar who

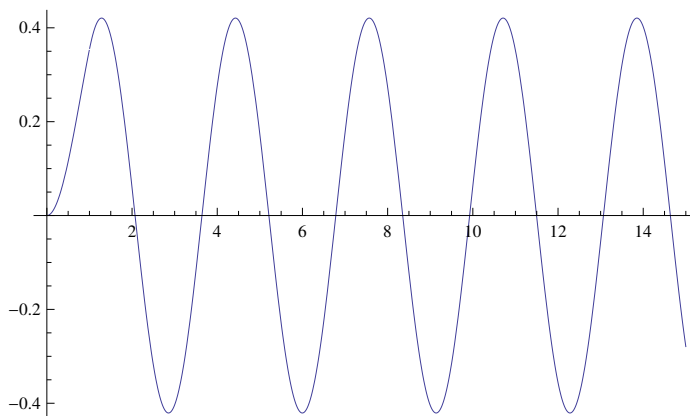


had been able to clearly demonstrate the bushy stunt virus had cubic symmetry [3, p 474].

Cubic symmetry relates to the rotational possibilities of the particle. It means that through the shell there are axes which one can rotate about  $120^\circ$  and the figure will end up in an identical position as it started. These axes are called 3-fold rotational axes. To see this in action, take a die and hold it by two opposing corners. Now rotate the die  $\frac{1}{3}$  of a turn. Although the position of the numbers has changed, the die is in an identical position as it started. Thus this die, a cube, is an example of a figure with cubic symmetry.

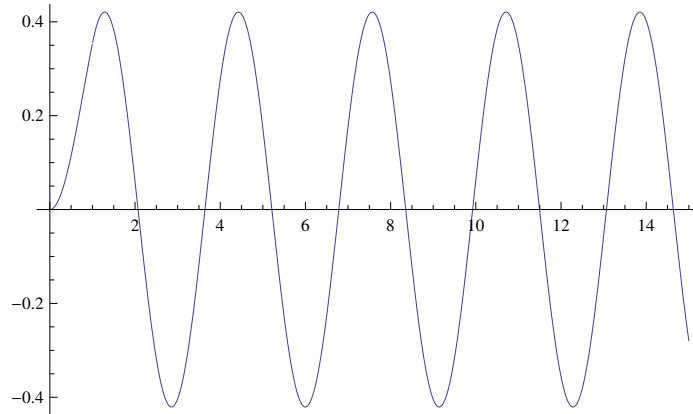
While the virus particle can be rotated, the individual sub-units cannot be "flipped". This is due to the presence of asymmetric carbon atoms [3, p 474]. This means there can be no reflective symmetry axes on the particle. For a visual representation of a figure with rotational symmetry axes but no reflective symmetry axes see Figure 3. Crick and Watson then examined all possible cubic point groups, which are listed in Table 4. [Is it worth defining a point group?]

Figure 3: Symmetry examples. Use figures from Mannige



They noticed that with each group, the number of possible asymmetric sub-units was a multiple of 12. For instance, with the 20 sided icosahedron, each side can be split into 3 asymmetric sub-units, thus creating a total of 60 which is of course a multiple of 12. Recall the reason they are concerned with the possible number of asymmetric units, not the number of faces, is because they wanted the smallest possible unit, and because the

Figure 4: Table of cubic point groups



individual proteins were bound to be asymmetric simply due to their nature. Additionally, this reasoning was influenced by their observation of the TMV particles, which had only protomeres. Thus, they were not considering the forms the proteins may aggregate into before the final shell was constructed, but instead concerned only with the individual units being in identical environments. This was their final, and ultimately erroneous conclusion, that the number of proteins in a viral shell will be a multiple of 12 [3, p 475].

It is important to note that their process was one of trying to apply a set of rules to the limited observations they had of both spherical and helical viruses. They were not trying to ask mathematical questions as to whether or not their results should be expected, but were instead simply trying to explain observed patterns. This technique, of describing observations, was also the route taken by Caspar and Klug a few years later, only with a more accurate outcome.

### 2.1.2 The Theory of Quasi-Equivalence

D. L. D. Caspar and A. Klug wrote many papers on the structure of viruses. Their most prominent work was done in 1962 in a paper entitled "Physical Principles in the Construction of Regular Viruses" [2]. In this work they built upon the ideas and groundwork set forth by Crick and Watson, only with more scientific tools at their disposal. The results were to

validate the general idea set forth by Crick and Watson, but to ultimately correct the theory and form what is still the basis of belief today.

Much of their paper repeats in greater detail the same concepts set forth by Crick and Watson. They both explained why there may only be a limited number of types of proteins in any one capsid and why the number of ways to efficiently design a container using a large number of identical proteins is very limited. Only Caspar and Klug had more scientific evidence, largely due to the greater volume of research that had been performed by 1962. For instance, by studying the weight of the RNA within small viruses, and the average number of nucleotides required to make a protein, they were able to prove that there is only enough genetic material in small viruses to produce 2 or 3 proteins [2, p 1]. This proved that with small viruses the number of types of proteins in the capsids is limited to 1 or 2. Additionally, advances in X-ray diffraction and electron microscopy had definitively shown for multiple small viruses that they were indeed made up of repeated identical sub-units [2, p 2]. What differed from Crick and Watson was the number of individual protomers possible in spherical viruses.

While Caspar and Klug ultimately disproved a flaw in Crick and Watson's theory for spherical viruses, it is worth noting that they also confirmed Crick and Watson's theory for helical ones. Caspar and Klug also relied heavily on the study of TMV for their understanding of helical viruses. Through advances in X-ray diffraction and other physical and chemical studies, they were able to definitively show that each individual protein in a TMV capsid is structurally and chemically identical to the others. They also found that the proteins would form into the helical shell without the presence of the TMV RNA [2, p 4-5]. These pieces of evidence added to the observations that the TMV capsid consisted of repeated units around a helical symmetry axis.

The final important observation made of the helical viruses was that they were flexible. The rod like capsids could bend and flex without breaking. Recalling that the helical capsids

are made up only of protomeres, the individual proteins, this meant that the inter-protein bonds did not have to be perfectly identical. The bonds were allowed to vary slightly, or be quasi-equivalent [2, p 7]. This discovery helped lead to the theory of quasi-equivalent bonds which allows for scalable spherical capsids.

### 2.1.3 Recent Work

This section will really quickly talk about some of the research that is being done on capsids currently. Twarock [4] [5] is using group theory to help explain the irregular capsids. They are spherical but do not strictly adhere to icosahedral symmetry, they have shifts, or translations in them. Mannige [6] explained the stress placed on bonds due to being skewed shells, thus explaining why they are less common. Zlotnick [7] and Zandi [8] are examples of those working to understand not only the structure of the capsids, but also how the form, how they come together, thus allowing us to understand how they might be destroyed.

I do not want this section to be long, just a brief sentence or two on each person to highlight the types of work still occurring.

## 3 Mathematical Problems and Solutions

Recall that the one of the goals of this paper is to explore the fun mathematical problems that can be inspired by biology. Thus, in the following sections mathematical problems from two different areas of icosahedral capsid theory will be explored. Multiple solution techniques will be given for the problems with varying degrees of mathematical difficulty.

### 3.1 Triangulation Numbers

There are two problems which arise from Caspar and Klug's definition and claims about their T-number. First, Caspar and Klug claim that their T-number,  $T = h^2 + hk + k^2$ ,

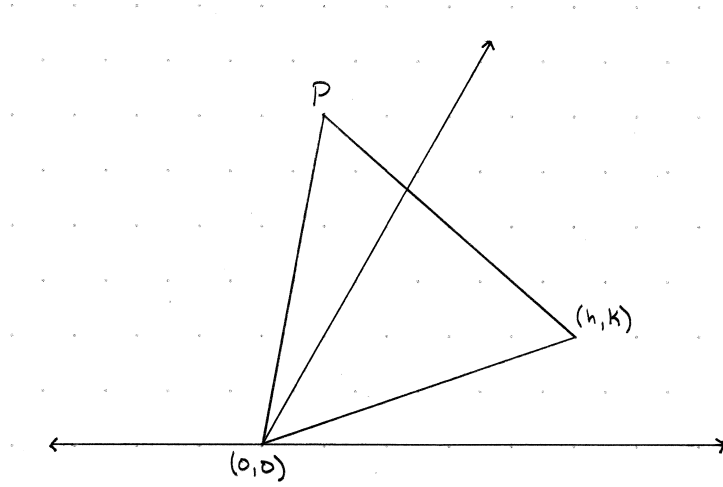
defines the exact number of identical asymmetric protein sub-units found in a spherical capsid through the formula  $P = 60T$ . Recall this 60 is because the T number measure the number of triangular facets per side of the icosahedron and each facet is composed of 3 protomeres, the protein sub-unit. Since an icosahedron has 12 sides, we would have  $3 * 12 * T = 60T$  proteins per capsid. While their T-number has been confirmed as accurate through better electron microscopes, the mathematical question remains, was this expected? Will the T-number always work for integers h and k. The first problem will be to establish that using points  $(0, 0)$  and  $(h, k)$  as the base of an equilateral triangle will result in a triangle will an area equal to T times the area of an equilateral triangle with unit length, that is base equal to one. The second question that must be asked is, if two corners are placed at  $(0, 0)$  and  $(h, k)$  will the third corner, call it point P, always fall on an integer valued point? (See Figure 5) To answer these questions we will take 2 different approaches. One requires no mathematics beyond high-school, and the second only requires a bit of knowledge of linear algebra.

### 3.1.1 Solution 1: Geometry

The goal of this approach is to establish that if we are given a point  $(h, k)$  as defined by Caspar and Klug (that is that the h-axis is the standard x-axis and the k-axis is a  $60^\circ$  rotation counter-clockwise of the h-axis) that the area of an equilateral triangle with a base from  $(0, 0)$  to  $(h, k)$  will equal the area of an equilateral triangle with base length 1 times  $T = h^2 + hk + k^2$ . This solution will be attempted with no more than the skills of someone who has passed high school geometry.

First, recall that the area of an equilateral triangle is  $A = \frac{\sqrt{3}}{4} b^2$  where  $b$  is the length of any side of the triangle. If this formula is unknown to the reader, it is a simple and beneficial exercise to re-establish it. To begin, take an equilateral triangle with side length  $b$  (Figure 6a). Draw a line from any corner to the midpoint of the opposing side (Figure 6b).

Figure 5: Equilateral Triangle on h, k-axes

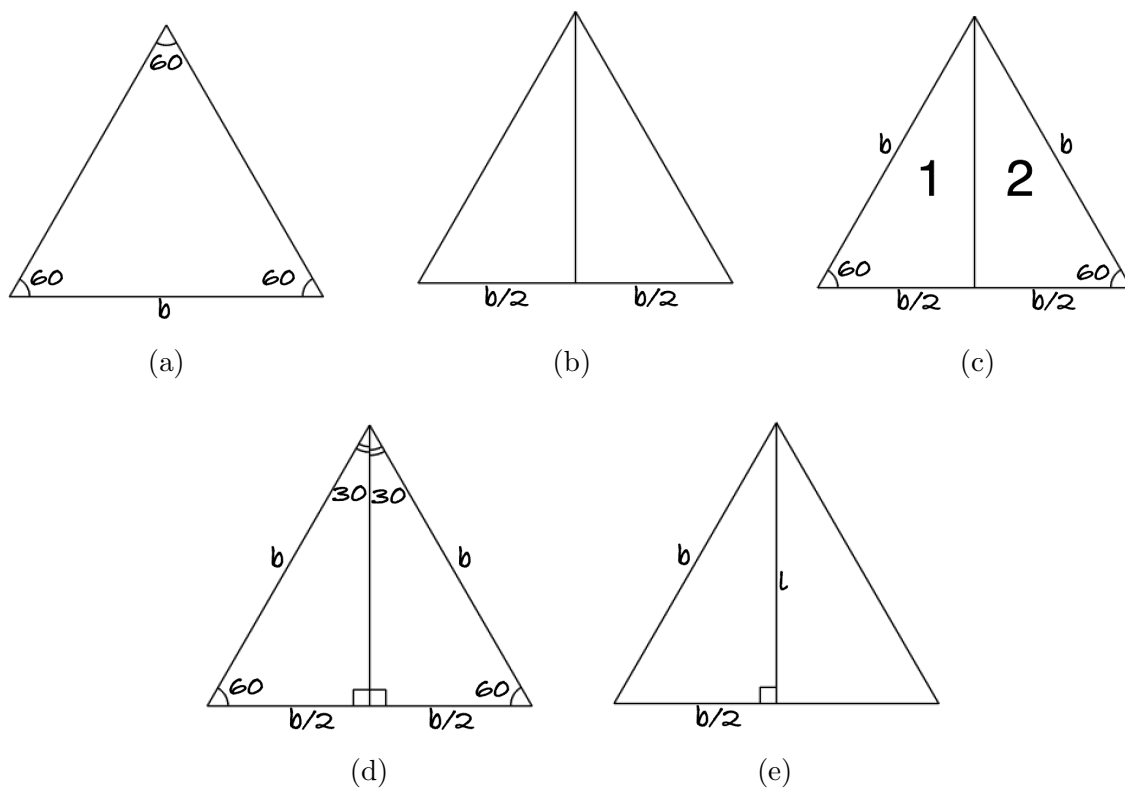


One face of a  $T = h^2 + hk + k^2$  icosahedral capsid is represented on the h, k coordinate system. Recall when working with T numbers that the h-axis is in the standard horizontal orientation, and the k-axis is at  $60^\circ$  to the h-axis in a counter clockwise rotation.

Now there are 2 triangles. Recalling the triangle congruency, by the side-angle-side theorem these triangles are congruent (Figure 6c). This forces the angle that was cut with the drawn line to be bisected. Since the starting triangle was an equilateral triangle which has all angles equal to  $60^\circ$ , the top angles are both  $30^\circ$  making the constructed line perpendicular to the base (Figure 6d).

Now to find the length of the constructed line  $l$  (Figure 6e). Thankfully the pythagorean

Figure 6: Area of Equilateral Triangle Constructions



theorem is all that is required for this.

$$b^2 = \left(\frac{b}{2}\right)^2 + l^2$$

$$l^2 = b^2 - \frac{b^2}{4}$$

$$l^2 = \frac{3}{4}b^2$$

$$l = \sqrt{\frac{3}{4}b^2}$$

$$l = \frac{\sqrt{3}}{2}b$$

Thus, using our standard triangle area formula of  $A = \frac{1}{2} \text{base} * \text{height}$  we get

$$A = \frac{1}{2} b * \frac{\sqrt{3}}{2} b$$

$$A = \frac{\sqrt{3}}{4} b^2$$

Recall the goal is to establish that the area of an equilateral triangle with a base from  $(0,0)$  to  $(h,k)$  (Figure 5) will equal the area of an equilateral triangle with base length 1 times  $T = h^2 + hk + k^2$ . So from the above formula with  $b = 1$ , the resulting area should be  $\frac{\sqrt{3}}{4} (h^2 + hk + k^2)$ .

Now consider the triangle created by drawing lines between  $(0,0)$ ,  $(h,k)$  and  $(h,0)$ . Two of the side lengths are known and side length  $b$  remains to be found (Figure 7). From the setup of the grid, the exterior angle at point  $(h,0)$  is  $60^\circ$ , so if a line is dropped down from point  $(h,k)$  perpendicular to the t-axis the result is a triangle that was just worked with and with known dimensions. Since it is  $k$  units from  $(h,0)$  to  $(h,k)$ , the perpendicular just drawn has length  $\frac{\sqrt{3}}{2} k$  and the final leg of the triangle is  $\frac{1}{2} k$  (Figure 8).

Figure 7

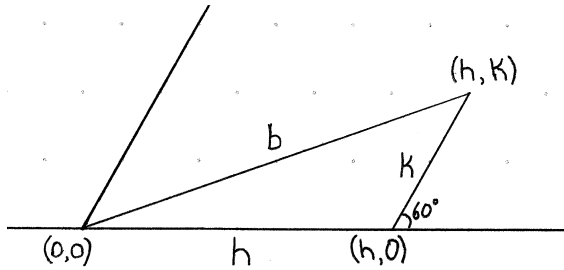
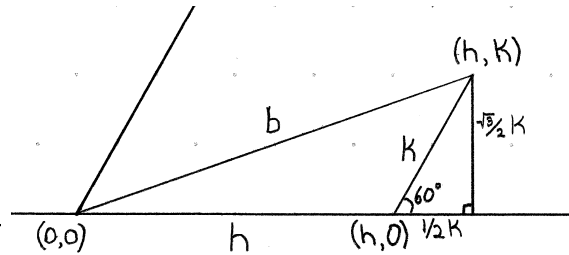


Figure 8



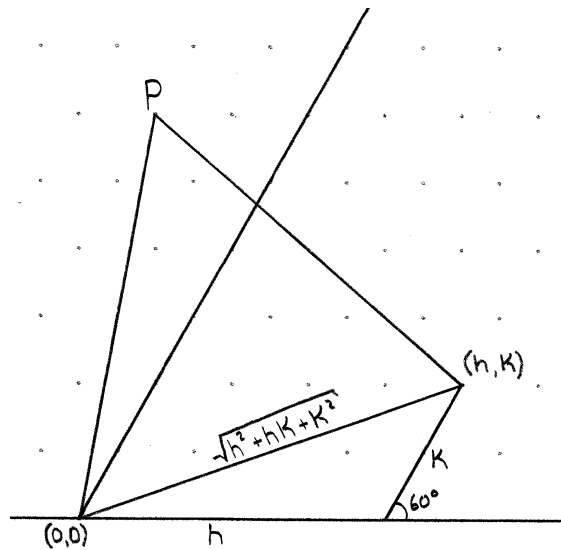
Now the length in question, from point  $(0,0)$  to  $(h,k)$ , is the hypotenuse of a right triangle with side lengths  $\frac{\sqrt{3}}{2} k$  and  $h + \frac{1}{2} k$ . Once again the solution lies at the end of the



pythagorean theorem.

$$\begin{aligned}
 L^2 &= \left( \frac{\sqrt{3}}{2} k \right)^2 + \left( h + \frac{1}{2} k \right)^2 \\
 &= \frac{3}{4} k^2 + h^2 + hk + \frac{1}{4} k^2 \\
 &= h^2 + hk + k^2 \\
 L &= \sqrt{h^2 + hk + k^2}
 \end{aligned}$$

Figure 9



Now plugging this into the Area formula for b, the result is  $A = \frac{\sqrt{3}}{4} (h^2 + hk + k^2)$ , which is exactly the desired area from before! So this definitively shows that the area of an equilateral triangle with base from point  $(h, 0)$  to  $(h, k)$  is equal to the area of  $T = h^2 + hk + k^2$  equilateral triangles with base length 1. Now recall:

1. The triangle with base from point  $(h, 0)$  to  $(h, k)$  represents one side of an icosahedron
2. Icosahedrons have 20 sides

3. The base 1 triangle represents one of the triangular facets of the capsid which is made up of 3 proteins

Put it all together and the result is that a total of  $3 * 20 * (h^2 + hk + k^2) = 60T$  proteins will fit on the capsid described.

**Solution 2:** Hey put cool math in here

### 3.1.2 Protein Shape

The second problem is not one that comes directly from the theory of quasi-equivalence or icosahedral symmetry. Instead, the second problem is a result of the standard images used in most papers on spherical viruses. When displaying capsids, most papers represent the protomers as trapezoids [ref figures I have used with trapezoidal subunits] with little to no explanation. The simplest explanation is that it is now possible to see the proteins in such clarity that they are known to resemble trapezoids. But again, this is bothersome. Should such a mathematically regular result have been expected and can it be explained? To state it another way, is there a mathematical justification for the protomer being represented by a trapezoid? Again two methods for answering this question will be provided. The first will again require no mathematical skills beyond that of a high-schooler, only some problem solving and algebraic manipulation of formulas. The second, while it can be viewed as a product of group theory, will boil down to nothing more than understanding symmetry.

### 3.1.3 Protein Shape

**This stuff does not need to be kept, i'm just not sure if I want some of it yet so I don't want to delete it** One constant and often unexplained attribute within works discussing icosahedral virus capsids is the authors' choice of subunit shape. Almost always

the subunit is represented by a trapezoid with little to no explanation of why. If we recap what we know so far:

1. Capsids are made up of pentamers and hexamers
2. Hexamers are capable of existing in a "flat" state
3. Three subunits need to make up an equilateral triangle

Given these conditions, especially point (3), we may simply guess that trapezoids are used because every equilateral triangle can be split into three identical (right word?) isosceles trapezoids (See INSERT figure).

But then we must ask why not some other shape that also has this property. For instance, equilateral triangles can also be split into three identical isosceles triangles or three identical right kites. To demonstrate, extend angle bisectors from each corner of an equilateral triangle until they meet in its interior. The result is that we have split the initial equilateral triangle into three identical isosceles triangles. Or we could create perpendicular bisectors on each side of an equilateral triangle and extend them within the interior of the triangle until they meet. This will result in the triangle being split into three identical right kites. (see INSERT figure) So again we must ask, why are trapezoids the standard subunit shape if other shapes will also meet our criteria of being able to fit perfectly inside an equilateral triangle?

Upon examination of shaded and deciphered images of real world icosahedral capsids we can see that the choice of trapezoid is simply due to their real world occurrence. (See [INSERT] capsid images) This was mentioned as early as (get date) by (get guy from mannige thesis) But this is a very interesting phenomenon. As pointed out by

**Solution 1:** More cool math using Euler's Polyhedral Formula (taught in early geometry classes). [9]

**Solution 2:** This one is kind of a group theory approach, but really just comes down to saying "Hey that shape is reflexively symmetric!" [?]

## 4 Conclusion

See it really wasn't that bad.

## 5 References

- [1] N. Campbell and J. B. Reece, *Biology*. San Francisco: Benjamin Cummings, 2002.
- [2] D. L. Caspar and A. Klug, “Physical principles in the construction of regular viruses.” *Cold Spring Harbor symposia on quantitative biology*, vol. 27, pp. 1–24, Jan. 1962.  
[Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/14019094>
- [3] F. H. C. Crick and J. D. Watson, “Structure of Small Viruses,” *Nature*, vol. 177, no. 4506, pp. 473–475, Mar. 1956. [Online]. Available: <http://dx.doi.org/10.1038/177473a0>
- [4] R. Twarock, “A tiling approach to virus capsid assembly explaining a structural puzzle in virology,” *Journal of Theoretical Biology*, vol. 226, no. 4, pp. 477–482, 2004.
- [5] —, “Mathematical virology: a novel approach to the structure and assembly of viruses.” *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, vol. 364, no. 1849, pp. 3357–3373, 2006.
- [6] R. V. Mannige and C. L. Brooks, “Geometric considerations in virus capsid size specificity, auxiliary requirements, and buckling,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 21, pp. 8531–8536, 2009. [Online]. Available: <http://www.pnas.org/content/106/21/8531.abstract>
- [7] A. Zlotnick, “Theoretical aspects of virus capsid assembly.” *Journal of Molecular Biology*, vol. 18, no. 6, pp. 479–490, 2005.
- [8] R. Zandi, D. Reguera, R. F. Bruinsma, W. M. Gelbart, J. Rudnick, and H. Reiss, “Origin of Icosahedral Symmetry in Viruses,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 44, pp. 15 556–15 560, 2004.  
[Online]. Available: <http://www.jstor.org/stable/3373675>

- [9] R. V. Mannige and C. L. Brooks, “Tilable nature of virus capsids and the role of topological constraints in natural capsid design,” *Phys. Rev. E*, vol. 77, no. 5, p. 51902, May 2008. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevE.77.051902>