Supplementary Problem I for Physics 6[1]

This problem is based on a famous 1956 paper by James Watson and Francis Crick (the discoverers of DNA). Starting with a few known facts from biochemistry, Watson and Crick applied simple geometry and some order-of-magnitude analysis to make a definite, detailed prediction (later found to be correct) about the chemical structure of viruses. Their calculation is a great example of how the estimation techniques we're learning can be applied to a problem in another field.

Things to remember while doing this problem:

- Don't be intimidated because it looks long! The problem has been broken down into many parts to lead you through the calculation in a step-by-step way; most of the answers can be found by putting together answers from the previous parts.
- Remember that this is an *estimate*: none of your answers should have more than one sig fig.
- The following units will be useful: 1 nm (nanometer) $= 10^{-9}$ m; 1 u (atomic mass unit) $\approx 10^{-27}$ kg. A typical atom has a mass of order 10 u.

At the time, it was known that a virus consists of a protein shell, just one molecule thick, with nucleic acid (genetic material) inside. The shape of this shell is essentially a sphere, with a diameter of around 40 nm. Watson and Crick were interested in finding out more about the structure of the protein shell: was it just one big protein molecule, or many smaller proteins joined together?

This was long before the genetic code had been worked out or any techniques for genome sequencing existed. Nevertheless, Watson and Crick were able to answer their question by thinking about the genes of a virus. They used the following key idea: **the size of a gene is proportional to the size of the protein it produces.** So if the shell of a virus were made of one gigantic protein, it would need a gigantic gene to code for this protein. Watson and Crick's goal (and ours, in this problem) was to figure out if such a large gene could fit inside a virus, given its known size.

a) Estimate the surface area of a virus's protein shell, in nm^2.

A typical "small" protein molecule is made up of a few thousand atoms, and so has a mass of about 30,000 u, or 30 ku. We'd like to estimate the surface area of such a molecule. Proteins typically curl up into very complicated shapes, but since this is a rough estimate, we can get a reasonable result by imagining the atoms simply arranged in a square.

b) If a protein molecule contains around 3000 atoms, about how many atoms should be on each side of such a square?

c) Given that the diameter of an atom is around 0.1 nm, estimate the surface area of this square, in nm^2. (Your result is probably a little large, since atoms tend to clump together and overlap when they form a molecule. Round it to the nearest order of magnitude.)

d) How many of these small protein molecules would be needed to make up the surface of a virus's protein shell?

CONTINUED ON BACK

[1] adapted from source material in William M. Gelbart and Charles M. Knobler, "The Physics of Phages," *Physics Today* **61**, 42 (2008).

e) If instead the shell consisted of just one large protein molecule, what would be its mass, in u?

Proteins are made up of long chains of amino acids; each amino acid is made up of around 10 atoms, and so has a mass of about 100 u.

f) Estimate the number of amino acids in the "small" and "large" protein molecules defined in parts (b) and (e).

In a gene, each amino acid is represented by one codon, which is made up of three nucleotides. (Nucleotides are the basic units that make up DNA, often represented by A, T, C, and G.) This is why larger proteins, with more amino acids, require larger genes. Watson and Crick knew that the mass of one nucleotide is about 330 u.

g) Estimate the mass of the genes needed to encode the "small" and "large" proteins.

Now we can figure out whether each of these genes would fit inside a virus. We'll assume (reasonably) that the nucleic acid has the same density as water: 1 g/cm^3, or 1000 kg/m^3.

h) Estimate the **volume** of nucleic acid that would fit inside a virus's protein shell, in nm^3.

i) Estimate the **mass** of nucleic acid that would fit inside a virus's protein shell, assuming it has the density of water. (Be careful of units!)

j) On the basis of your results, is it possible that the protein shell consists of one large protein? Explain why or why not.