LESSON 7: INSIDE A CAPSID, and CLASSIFYING VIRUSES

We've already learned the basic parts of a virus, but now let's take a closer look at the inside parts. We'll have to be very general in our descriptions because no virus is exactly like any other.

Inside a virus we find basically two things: 1) a genome in the form of either DNA or RNA, and 2) one or more tiny protein "tools" that the virus will need upon entering the cell. Some viruses, like polio, are too small to carry extra proteins and will borrow or make everything once inside a cell. Other viruses, like smallpox, are very large and have room to carry over a hundred protein tools inside their capsid.

Genome length is measured in base pairs (rungs of the DNA or RNA ladder). Parvovirus has one of the smallest genomes, at less than 3,000 base pairs. Polio is also fairly small, at about 8,000. Coronavirus has one of the largest RNA genomes at almost 30,000 base pairs. When genomes are larger than about 10,000 base pairs, they begin to be fragile. You will find larger genomes wound around protein strings, rods, or spools, to protect them from damage. The entire structure (genome plus proteins) is called the **nucleocapsid**.

Genomes can be either DNA or RNA. For DNA genomes, it can be the usual double-stranded DNA (the same stuff found in the nucleus of all cells), or it can be single-stranded DNA (a ladder with just one side).

-- Double-stranded (ds) DNA viruses include many we are familiar with such as adenovirus, papilloma (warts), all the herpes viruses, and the pox family of viruses. Viruses with dsDNA can use the host cell's copying machines in the nucleus.

-- Single-stranded (ss) DNA viruses are on the small side and include parvo, circo, and anello, and many plant viruses. When a single-stranded viral genome comes into the nucleus of a host cell, the cell's DNA machinery thinks that the viral DNA needs to be repaired, and it will go to work making a matching strand, and thus turning it into dsDNA, which is great for the virus because all the cell's machine are made for dsDNA.

--Another variation on DNA is the "gapped" DNA viruses. Their DNA is generally circular, with most of it double-stranded, but gaps were it is only single stranded. The cell machinery sees the gaps and "fixes" them, thus allowing the viral DNA to be able to use the cellular copying machines. Examples of gapped DNA viruses included Hepatitis C and feline (cat) leukemia.

RNA genomes can be either "positive sense" or "negative sense." Both DNA and RNA are directional, meaning that it makes a difference which end you start from. It's like having a right side and a left side, except that right and left doesn't make sense. The two ends are called 5' (five prime) and 3' (three prime). These numbers come from the arrangement of the 5 carbon atoms in the ribose sugar found in the "backbone" (sides of the ladder). Scientists decided that RNA that can be fed into ribosomes (mRNA) would be called "positive sense." (Ribosomes read RNA starting with the 5' end.) RNA that is read the other direction, staring with the 3' end would be called "negative sense." When "positive" and "negative" are used in this way, there is no electrical meaning at all. It's simply a way to keep track of the directionality of the molecule.

--Positive sense RNA viruses are the only viruses whose genome can be read by ribosomes. RNA(+) viruses include the caliciviruses (the ones the whales poop out), coronaviruses, flaviviruses (yellow fever), and the Picorna family (members include polio, rhinovirus (causes colds), and Coxsackie B).

--Negative sense RNA viruses are just as successful as positive sense ones, despite the fact that they have to have a more complicated system for making mRNA to feed to ribosomes. They have to carry the code for making a copying machine that uses RNA to make RNA. Cells don't have this type of copy machine. Some RNA(-) viruses not only carry the code in their genome, they also carry the actual protein machine inside their capsid. Influenza has an RNA polymerase protein stuck onto the ends of each one of its 8 RNA segments.

--Some viral genomes have doubled-stranded RNA, looking very much like DNA, but having A, U, C and G, as their bases, instead of the ones found in DNA: A, T, C, and G. The Reovirus family has dsRNA. If you've ever had a "stomach flu" you might have had an encounter with a rotavirus, a member of this group.

--The retroviruses have RNA(+) but are in a group all to themselves because they have a very different strategy. They have a special copying machine called "reverse transcriptase" which lets them make DNA from their RNA. They then insert this DNA copy into the genome of the host cell, so it becomes a permanent part of the cell's genome. The most well-known member of this group is HIV (Human Immunodeficiency Virus). Reverse transcriptase was discovered by David Baltimore, the same person who invented this classification system for viruses.

Genomes are very orderly in their arrangement of information. At one end you find all the coding for the structures such as capsid, spikes, or matrix. At the other end you find the coding for how to build copy machines and other tools (scissors, staplers, folders, etc.). The purpose of some tools is to damage cellular machines. In many viral genomes you also find codes that tell when certain proteins are to be made or used. The genome is generally read from the 3' prime first, where all the tool information is. The structural proteins come last.

There are key items that all viruses need but <u>none</u> of them have: **ribosomes, raw materials, and a source of energy**. All viruses must use the cell's ribosomes to make their protein structures from the cell's supply of amino acids, and all viruses must rely on the cell's production of ATP energy.