



Let's Get Something Straight

by **Charles H. Calisher**
calisher@cybercell.net

> **Croat Med J. 2007;48:103-6**

Sequences vs Viruses: Producer vs Product, Cause and Effect

In 1859, Charles Robert Darwin published "The origin of species." Six years later, Gregor Johann Mendel, an Augustinian abbot, published "Experiments on plant hybridization," a summary of his studies of inherited traits in pea plants. These two fundamental intellectual achievements have led to a remarkable understanding of inheritance, that could not have been dreamed of 200 years ago. Darwin, the naturalist, was overwhelmed by what he saw as evidence that species originated via evolutionary change and proposed the theory that natural selection was the mechanism by which such changes occur.

In 1869, Friedrich Miescher discovered "nuclein" and later isolated and purified this substance from salmon sperm. Twenty years later, his pupil Richard Alt-

mann, named this substance "nucleic acid" and showed that it occurred only in chromosomes, the basic hereditary unit. Shortly after, the classical term "genetics" was invented by William Bateson. By the 19th century, biochemists had isolated deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) from cell nuclei. It was realized only later that nucleotides were of two types – DNA containing deoxyribose and RNA containing ribose. In 1910, Thomas Hunt Morgan showed that genes resided on chromosomes and discovered linked genes on chromosomes. Then Alfred Sturtevant mapped a chromosome and others showed that genes are linearly arrayed on chromosomes.

By 1927, the word "mutation" had been coined for physi-

cal changes in genes and Frederick Griffith discovered the next year that hereditary information, later shown to be DNA, could be transferred between bacteria. The rapidity with which discoveries were made in the field of genetics was becoming astonishing:

- 1929: Phoebus Levene identified the components of DNA (the four bases, the sugar and the phosphate chain) and showed that these components were linked in the order phosphate-sugar-base. He called each of these units a nucleotide and suggested that the DNA molecule consisted of a string of nucleotide units linked together through the phosphate groups, the "backbone" of the molecule. However, Levene thought the chain was short and that the bases repeated in the same fixed order, but Torbjorn

Caspersson and Einar Hammersten showed that DNA was a polymer;

- 1931: Barbara McClintock demonstrated that crossing over was the cause of recombination;

- 1941: Edward Tatum and George Wells showed that genes coded for proteins;

- 1944: Oswald Avery, Colin McLeod, and Maclyn McCarty isolated DNA as the genetic material;

- 1950 Erwin Chargoff discovered that the four nucleotides (adenine, thymine, cytosine, and guanine) that comprise DNA were not found in constant proportions but that adenine and thymine always remained in equal proportions, as do cytosine and thymine, and Barbara McClintock discovered transposons in corn (maize);

- 1953, Alfred Hershey and Martha Chase, using a bacteriophage, proved that DNA was the genetic material.

Another great discovery was made in 1953. With fundamental clues provided by Rosalind Franklin, James Watson and Francis Crick demonstrated that the structure of the DNA molecule was that of a double helix, stating that this "suggests a possible copying mechanism for the genetic material." In that same issue of *Nature*, Maurice Wilkins and colleagues reported their analyses of x-ray crystallographic studies, which suggested that the structure existed in biological systems.

Based on all these findings, numerous discoveries followed. It was shown that the correct chromosome number for humans was 46 (not 48), that the genetic code was arranged in triplets, that restriction enzymes could be used to cut and paste sections of DNA, and then, in 1983, Kary Mullis devised the extraordinary polymerase chain reaction, enabling relative ease of nucleic acid amplification. That seminal development provided a rapid means of analyzing sequences and within the next 20 years a human gene had been sequenced and 99% of the human genome had been sequenced with 99.99% accuracy. Many of these investigators, and others who made fundamental contributions to our understanding of the chemical structures (genotypes) and biological activities (phenotypes), were awarded Nobel Prizes.

The conceptual difficulty at the time was in grasping how specific (identifiable) characteristics of the structure of DNA molecules could lead to specific traits and behaviors in the organism harbouring those molecules. After all, chemical analyses of DNA always showed the presence of the same four nucleotides and the structure of the DNA molecule appeared to be reasonably simple and uniform. In contrast, the organism always appeared to be remarkably complex and organisms in toto are widely diverse. After the discov-

ery of the structure of DNA, it became apparent that segments of the nucleic acid polymer (the genes) were responsible for specific features. We now know that the genetic code can be equated to a map of the information coded for by the particular nucleic acid sequence. The sequence specifies amino acid products and those amino acids become part of various proteins, essential components of all organisms. These gene products include enzymes, structural elements, those that participate in cell signaling, in the immune response, and, in effect, all other functions of cells.

Finally, proteins are amino acid polymers whose sequences are specified by the genes encoded by the nucleic acid sequence (the genome). First, the gene produces a protein known as RNA polymerase. This enzyme reads (transcribes) the nucleotide sequence and synthesizes a messenger RNA molecule. After transcription of the message, ribosomes in the cell read (translate) 3-base segments (codons) of the messenger RNA and add amino acids as specified by the genes.

All this is remarkable. Remarkable that the mechanisms occur as they do, remarkable that we understand as much as we do, and remarkable how much more there is for us to learn and understand. But how does cellular DNA and cellular

RNA relate to viral DNA and viral RNA?

In general, viruses comprise a genome within a protective protein coat or shell, a capsid. Viruses do not multiply, they replicate. More correctly, they are replicated by the cell mechanisms, which are subverted by the infecting virus and under the direction of that virus. By themselves, viruses do nothing, except to exist as packages of nucleic acid and protein. Viruses vary in shape, size, nucleic acid type (DNA or RNA; some have both), whether they can infect and affect plant cells, vertebrate cells, invertebrate cells, cells of bacteria, cells of parasites, cells of fungi, or other types of cells. Viruses may have specificity for cells of the liver, lungs, intestines, or heart; of bird brains, kangaroo eyes, mosquito ovaries, rodent kidneys, salmon pancreases; and of numerous other life forms. Depending on the virus, they may cause pneumonias, renal failure, blindness, rash, arthralgias, encephalitis and other neurologic disorders, common colds, smallpox, hemorrhages, immune deficiency, skin lesions, sterility, cervical carcinoma, other malignancies, and many other illnesses. They may be responsible for multiple sclerosis and chronic fatigue syndrome; and they surely will be shown to cause other diseases. Viruses may affect the host by causing fatal infections, merely make the host ill (very ill or mildly ill), or infect

the host and not cause any obvious illness. Viruses that infect hedgehogs (insectivores) may or may not infect birds, or viruses that infect hedgehogs but are not pathogenic for them may be pathogenic for birds; those that infect certain rodents may simply replicate in them and be shed in their urine but when they infect humans they may cause fatal illnesses of the kidneys or respiratory tract; viruses that are pathogenic for humans may replicate only in humans and other primates; viruses may not replicate in mosquito cells or may replicate to high titer without causing cytopathic (degenerative) effects; they may replicate only in beetles; may be pathogenic for bees or trees or peas, and maybe be carried on a breeze, or a sneeze, if you please.

The point here is that the many viruses (71 families, containing 282 genera, incorporating many thousands of viruses) are widely diverse. Their characteristics and properties are commensurately widely diverse, and there is job security for all virologists for many years (if we can obtain funding).

One huge confusion about taxonomy is that the term taxon equals the term virus. It does not. A virus taxon is a ranked grouping of viruses forming a hierarchy, a list, like a laundry list. The word "bread" on a laundry list does not contain flour and it is not edible. The word "bread" is not bread; it is a symbol for

the physical substance we enjoy with butter, cheese, meat or jam. One cannot eat a symbol. Likewise, one cannot be infected by the word "virus" or acquire an illness from the term "Severe acute respiratory syndrome" and those words do not replicate in bats. It is the virus, not the word "virus," that is infectious and potentially dangerous. In the same way, a virus sequence is not a virus, any more than a leg is a person.

Using viruses as an example (although the same analogy could be made for mammals, birds, fish, plants, and politicians, all of which have genomes), the sequence of the virus genome is the specifier, the conductor of the cellular orchestra, the anti-democratic, authoritarian, and despotic tyrant that subverts the cellular proclivities, the dictator of the replication cycle. Strong words, indeed, but true ones. Without viral subversion, the cell goes about the business of whatever cells do. In the presence of a viral nucleic acid sequence, it is simply a matter of time before the cell commits suicide. Still, and to the point of this diatribe, the sequence is not a virus. The virus contains a nucleotide sequence, or sequences, but it has other components, with which to attach to cell surfaces, agglutinate erythrocytes, arouse the immune system, and it may have a multitude of other properties.

Even when we know the entire nucleic acid sequence of the virus, we do not know all of its properties. Certainly we can identify its relatives (flaviviruses, herpesviruses, retroviruses, caulimoviruses, orthomyxoviruses, or others) and perhaps deduce some of its properties, such as replication in arthropods, infections of humans, and lesion-production in certain plants, but that is not guaranteed by any means. If we know the entire nucleic acid sequence of a newly discovered virus do we know its epidemic potential? No. The phenotypic properties of a virus are reflections of properties specified by the sequence but they are not the sequence itself. Complicating matters, the hosts may not respond uniformly, because of genetic variations among individuals, the presence of antibodies (homologous or heterologous to the infecting virus), concurrent infections with

interfering viruses, age, gender, genetic polymorphism of enzymes, nutritional factors, histocompatibility antigens, protection by drugs or by natural products, and so on. The fact is that there is no way to predict the exact phenotype from having knowledge of the genotype.

I suggest to viral molecular biologists that they devote more time to the study of the biology of viruses, so that the glorious data they produce has greater biological meaning than it now has; otherwise, they should change their title to "molecular chemist". As for those of you reading this who are ignorant of the near-miracles being wrought by molecular means, do not smirk and think I am agreeing with your biases. Instead, go find a young person and invite her or him for lunch and a chat. The camouflaged plea here is for the virologic and bacteriologic communities to isolate, char-

acterize and store as museum specimens, viruses and bacteria. Simply knowing their sequences is cold, unrewarding, and ultimately short-sighted. There are enough people now who think that sequences and viruses come from freezers.

Further reading

- Calisher CH, Mahy BW. Taxonomy: get it right or leave it alone. *Am J Trop Med Hyg.* 2003;68:505-6
- Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA, editors. *Virus taxonomy: eighth report of the International Committee on Taxonomy of Viruses.* Amsterdam: Elsevier Academic Press; 2005.
- James Watson, Francis Crick, Maurice Wilkins, Rosalind Franklin (overview). Available from: <http://www.cbeheritage.org/classroom/chemach/pharmaceuticals/watson-crick.html>. Accessed: January 8, 2006.
- Maurice Wilkins. Available from: http://nobelprize.org/nobel_prizes/medicine/laureates/1962/wilkins-bio.html. Accessed: January 8, 2006.
- Watson JD, Crick FH. A structure for deoxyribose nucleic acid. *Nature.* 1953; 171:737-8.
- Wilkins MH, Stokes AR, Wilson HR. Molecular structure of deoxypentose nucleic acids. *Nature.* 1953;171:738-40.