Frugal mechanisms of viral genome expression

Clever Coding

ADRIANNA ŁONIEWSKA-ŁWOWSKA
Institute of Biochemistry and Biophysics, Warsaw
Polish Academy of Sciences
adalon@ibb.waw.pl

Dr. Adrianna Łoniewska-Łwowska works on obtaining an infectious copy of the potato virus Y (PVY) genome and on identifying the regulatory elements involved in the synthesis and expression of proteins from subgenomic RNA

Although viral genomes are very small, ingenious mechanisms now being discovered enable them to encode a sufficient number of proteins

Every one of us has encountered viruses, at the very least in the form of an unpleasant nose cold. But unfortunately, viral activity is not limited just to such non-life-threatening illnesses in humans. Such serious diseases as AIDS, viral hepatitis, and rabies are also caused by viruses. Plants are likewise attacked by certain viruses, such as the tobacco mosaic virus (TMV), the first virus ever discovered, or potato virus Y (PVY). Viral infections can significantly affect yields of important crops. There are also some viruses that attack bacteria, called bacteriophages - structurally very complex viruses discovered in the early 19th century by Canadian doctor Félix Hubert d'Herelle.

It seems that there is no living organism that is unaffected by viruses. It is a little known fact that under the structural definition of life, these menacing parasites are actually not considered to be living organisms themselves. They have no cell structure and they are incapable of growth or reproduction outside a host cell. Yet on the other hand,

Jack Dykinga, ARS Image Gallery

A leaf attacked by the tobacco mosaic virus - the first virus ever discovered

they do have genes and demonstrate a capacity to evolve.

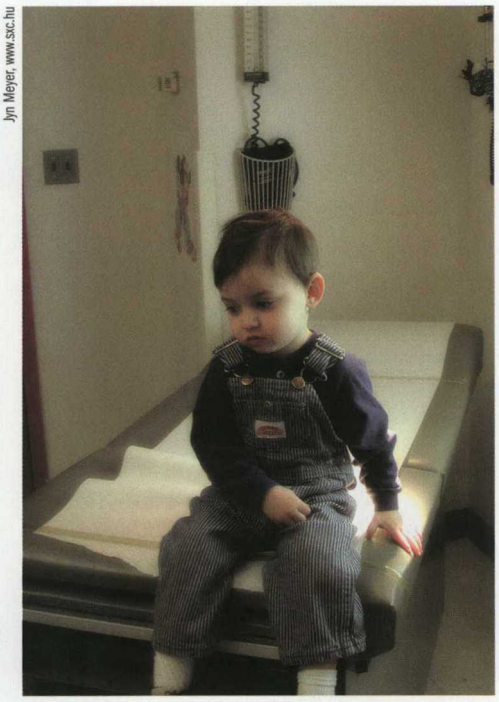
Ever-present pests

Humankind has always had to cope with viruses. The first reports describing a human being with symptoms typical of poliomyelitis, a viral disease, date to before ca. 1400 BC. We also know that smallpox was an endemic disease in China around 1000 BC. The first attempts at vaccination were even made in those times. However, quite some time would have to pass before humans could study and learn to fight these invisible predators. It was not until the 1940s that the invention of the electron microscope made it possible to directly observe virus particles and to evaluate their size (until then measured in terms of permeability through a bacteria filter), their quantities within the material studied, and their structure.

An individual virus particle, called a virion, consists of two elements. One of them is the genome - the infectious portion in the strict sense - in the form of DNA (deoxyribonucleic acid) or RNA (ribonucleic acid). The other element is a surrounding protein shell, also known as a coat or capsid, which serves to protect the genetic material inside. Some viruses moreover have an envelope of glycoproteins or lipids on the surface of the capsid, which help recognize the cell to be infected and facilitate the fusion of the capsid tissue with the cell membrane.

Based on the spatial structure of the capsid, we can classify virons into two basic types: spherical (or icosahedral, such as the polio virus) and cylindrical (or helical, such as TMV). Interestingly, if we mix nucleic acids with capsomeres (the compounds comprising the capsid), new virons will actually assemble themselves spontaneously, yielding infectious forms. Moreover, in the absence of genetic material, capsomeres themselves can independently form what are called viruslike particles (VLP). Such particles cannot cause any viral infection because they do not contain any genetic material, yet they are capable of triggering the production of antibodies within the body, so they can be used in producing vaccinations. The most complex morphology is seen in viruses which attack bacteria (known as bacteriophages). They consist of a head encapsulating the genetic material plus a tail with a contracting sheath, sometimes with additional tail fibers. By contracting, the tail breaks the bacteria cell wall and injects the genetic material into the cell.

Jyn Meyer, www.ssc.hu



Viruses have been around since the dawn of mankind. Most childhood illnesses are in fact of viral origin

Scanty genomes

A breakthrough in humanity's battle against viruses came with the rise of molecular biology. The development of the techniques of DNA recombination, sequencing, thermocyclic polymerase chain reaction (PCR), restrictive analysis, and especially the technique of "reverse genetics" have allowed the structure of viral genes and genomes to be identified. That, in turn, has enabled us to answer the question of how genes are responsible for the "act of viral violence."

In-depth study has shown that viruses have genomes that are relatively small, compared to those of prokaryotic cells (bacteria) or eukaryotic cells (nuclear organisms: plants and animals). The largest known viral genome contains only 190 thousand nucleotides (190 kb), while bacterial genomes are on the order of 1000-9000 kb long. The human genome, in turn, consists of as many as 6 billion nucleotides! Because of their small size, it is obvious that viral genomes can only code a limited number of proteins.

PHOTOTAKE/BE&W

A breakthrough in humanity's battle against viruses came with the rise of molecular biology

Depending on the type of virus, that ranges from five up to several hundred proteins. Some of them are used to build the capsid. Viral genomes also code their own DNA and RNA polymerases, meaning enzymatic proteins which produce daughter copies of the viral genome. Those proteins are necessary because the host cells do not contain either the RNA polymerases necessary for replicating viral genomes made of RNA, or the necessary DNA polymerases and additional proteins required for replicating certain viral DNA. Some viral genomes likewise code the proteins responsible for generating pathological symptoms or for determinants of their host-to-host transport via "vectors," such as insects.

Genetic economy

Even though viral genomes do not code many genes (as compared to pro- and eukaryotic organisms), it was still puzzling how even such numbers of genes could fit within such small genomes. Two important traits of viral genomes turn out to be responsible for their high concentration of genetic information. Firstly, a vast majority of the nucleotides in viral genomes are actually involved in the coding of proteins, which means that a viral genome contains very few non-coding sequences. This is the complete opposite of what is seen in many eukaryotic organisms, including humans, where a significant percentage of nucleotide sequences in the genome are non-coding.

Secondly, many viruses exhibit a kind of twofold utilization of genetic information, rarely encountered in eukaryotic organisms. This can significantly boost the coding potential of the viral genome, above what we might expect based on the genetic code itself. But how can a single sequence encode more than one protein? That question has been answered by years of research, which discovered that viruses employ diverse mechanisms of expression for translating nucleotide sequences into the amino acid sequences of proteins. In general, sequences which encode genes have "start" and "stop" signals, telling the cellular protein-building machine where to begin and end reading. Sometimes there are several "start" signals close to one another in the sequence. Under the accepted model of classic translation, initiation should begin from the first "start" signal, but sometimes the translation machine happens to synthesize proteins from both the first and second "start" signals. As a result, two pools of viral proteins can be created, with partially the same or completely

different amino acid sequences. This mechanism is known as "leaky scanning" and it is utilized by the Human Immunodeficiency Virus (HIV) and the Human Papilloma Virus (HPV), to name examples.

Running through a red light

Another frequently employed mechanism of expression which only occurs under favorable conditions is called a "read through." Here it is as if the translation machinery failed to stop at a red light and runs straight through. But rather than causing a crash, so to speak, instead of a protein with a shorter sequence the process simply yields a different protein with a longer, partially identical sequence.

One more mechanism viruses use to boost the efficient harnessing of their scanty genome produces a change in the reading frame. That causes the very same nucleotide sequence to be utilized twice, albeit in a different phase. After at the point where the reading frame is altered, the two emerging proteins may evidence completely different amino acid sequences.

Sometimes the structure of the genome essentially forces viruses to synthesize additional RNA classes (called subgenomic RNA). That means that genes which were not accessible to the translation machine in the genomic arrangement are now effectively expressed. This is a phenomenon encountered in many viruses of both plants and animals.

The occurrence of any of these mechanisms depends upon the sequences that regulate virus processes. These may be structural or sequencing elements (such as promoters) that arise within the genome.

The potato leafroll virus

The present author's own research focuses on the potato leafroll virus (PLRV). This is a plant pathogen with a relatively small genome (around 6,000 nucleotides) in the form of single-strand (+) RNA, within which 8 genes have been distinguished, expressing functional and structural proteins. The former are involved in the replication of the virus (polymerase) or its movement within the plant, known as "movement proteins" (MP), while the latter build the capsid, known as "coat proteins" (CP).

To code these proteins, this specific virus employs many of the expression mechanisms outlined above. However, the location and nature of the regulator sequences long remained unidentified. My experiments managed to pinpoint both the promoter sequence of subgenomic RNA1 and the region responsible for synthesizing proteins of this particle. Both regulatory regions are located in the vicinity of the protein-coding sequence, which means that two functions are superimposed within a single sequence: regulatory and coding functions. This of course paves the way for further discussion about the co-evolution of sequences and the demands nature makes of viral genomes. But one thing is certain: PLRV serves as an excellent example of the economy principle of genetic information, making ideal use of what nature has given it – a scanty genome. ■

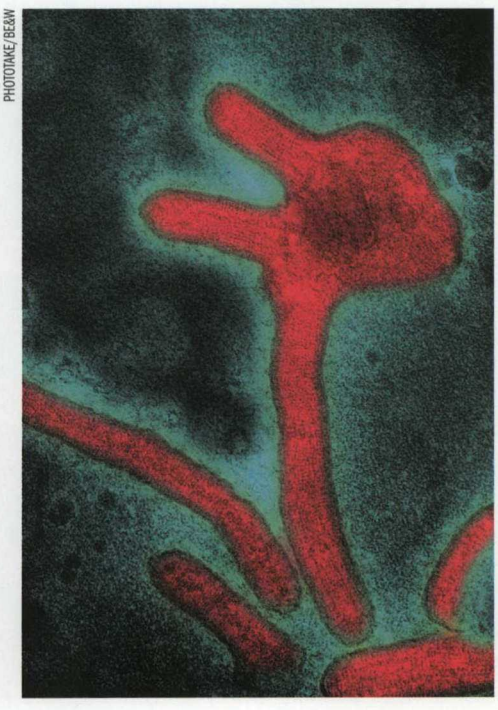
Further reading:

Gale M.Jr., Tan S.L., Katze M.G. (2000). Translational control of viral gene expression in eukaryotes. *Microbiol Mol Biol Rev., 64 (2)*, 239-80.

Bernardi F., Haenni A.L. (1998). Viruses: exquisite models for cell strategies. *Biochimie, 80 (12)*, 1035-41.

Maia I.G., Séron K., Haenni A.L., Bernardi F. (1996). Gene expression from viral RNA genomes. *Plant Mol Biol, 32 (1-2)*, 367-91.

PHOTOTAKE/BE&W



Viruses employ various techniques to evade the problem of their small genome. Here a Marburg virus, which causes hemorrhagic fever